



# Quelques contributions à l'estimation fonctionnelle par méthodes d'ondelettes

Christophe Chesneau

## ► To cite this version:

Christophe Chesneau. Quelques contributions à l'estimation fonctionnelle par méthodes d'ondelettes. Mathématiques [math]. Université Pierre et Marie Curie - Paris VI, 2006. Français. NNT: . tel-00121364

**HAL Id: tel-00121364**

**<https://theses.hal.science/tel-00121364>**

Submitted on 20 Dec 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ PIERRE ET MARIE CURIE - PARIS VI

# THÈSE

pour obtenir le titre de

DOCTEUR EN SCIENCES

DE L'UNIVERSITÉ PIERRE ET MARIE CURIE - PARIS VI

Spécialité MATHÉMATIQUES

présentée et soutenue par

Christophe CHESNEAU

---

## Quelques contributions à l'estimation fonctionnelle par méthodes d'ondelettes

---

Thèse dirigée par Gérard KERKYACHARIAN

Présentée publiquement le 07 décembre 2006  
devant le jury composé de

M. Gérard KERKYACHARIAN	DIRECTEUR DE THÈSE
M. Anatoli JUDITSKY	RAPPORTEUR
M. Laurent CAVALIER	RAPPORTEUR
M. Alexandre TSYBAKOV	EXAMINATEUR
M. Stéphane BOUCHERON	EXAMINATEUR



# Remerciements

Tout d’abord, je tiens à remercier chaleureusement Gérard Kerkyacharian pour ses conseils éclairés, son soutien continu, ses qualités humaines rares et l’expérience scientifique dont il m’a généreusement fait profiter.

Je tiens à exprimer ma plus profonde gratitude à Anatoli Juditsky et Laurent Cavalier pour avoir accepté de rapporter ma thèse et pour l’intérêt porté à mon travail.

Je suis très honoré que Anatoli Juditsky, Laurent Cavalier, Stéphane Boucheron, Alexandre Tsybakov et Gérard Kerkyacharian soient les membres de mon jury.

Mes remerciements s’adressent également à Dominique Picard, Vincent Rivoirard, Florent Autin et Erwann Le Pennec pour leur disponibilité et leur grande gentillesse.

Je salue les nombreux doctorants avec lesquels j’ai eu de longues discussions amicales et parfois passionnées. J’ai une pensée toute particulière pour Vathana Ly Vath, Thomas Willer, Guillaume Lecué, Claire Lacour, Olivier Wintenberger, Mathieu Rosenbaum, Stéphane Gaïffas, Pierre Alquier, Philippe Rigollet, Frédéric Guilloux, Tu Nguyen, Thanh Mai Pham Ngoc et Sébastien Darses.

Il est important pour moi de remercier ma famille qui a toujours été une source inépuisable d’encouragements.

Mes plus tendres pensées vont tout naturellement à Stéphanie pour sa patience et son soutien constant au cours de ces années.

Je souhaite enfin rendre hommage à mon grand-père Rénato, et ma tante Graziella, qui auraient sans aucun doute appréciés ce travail.

\* \* \*



# Table des matières

<b>1</b>	<b>Présentation</b>	<b>7</b>
1.1	Cadre mathématique . . . . .	7
1.2	Contenu de la thèse . . . . .	14
<b>2</b>	<b>Ondelettes et estimation fonctionnelle</b>	<b>21</b>
2.1	Introduction . . . . .	21
2.2	Bases d'ondelettes . . . . .	22
2.3	Estimateurs en ondelettes . . . . .	25
2.4	Premiers résultats . . . . .	27
<b>3</b>	<b>Vitesses minimax : boules de Besov</b>	<b>33</b>
3.1	Introduction . . . . .	33
3.2	Bornes supérieures et inférieures . . . . .	34
3.3	Bruit blanc gaussien généralisé . . . . .	37
3.4	Régression à pas aléatoires . . . . .	41
3.5	Démonstrations . . . . .	44
<b>4</b>	<b>Étude minimax : boules de Besov pondérées</b>	<b>59</b>
4.1	Introduction . . . . .	60
4.2	Outils mathématiques . . . . .	61
4.3	Bornes supérieures . . . . .	65
4.4	Applications . . . . .	67
4.5	Simulations graphiques : régression à pas aléatoires . . . . .	70
4.6	Démonstrations . . . . .	77
<b>5</b>	<b>Seuillage par blocs</b>	<b>83</b>
5.1	Introduction . . . . .	84
5.2	Estimateurs et hypothèses . . . . .	86
5.3	Encadrement du risque . . . . .	88
5.4	Résultats minimax . . . . .	90
5.5	Résultats maxisets . . . . .	94
5.6	Applications . . . . .	96

5.7 Démonstrations . . . . .	103
<b>6 Conclusions et perspectives</b>	<b>133</b>
<b>Références</b>	<b>135</b>

# Chapitre 1

## Présentation

Ce chapitre décrit le cadre mathématique et le contenu de la thèse.

### Sommaire

<b>1.1</b>	<b>Cadre mathématique</b>	<b>7</b>
1.1.1	Point de départ	7
1.1.2	Approche minimax	9
1.1.3	Approche maxiset	11
1.1.4	Particularités de notre étude	13
<b>1.2</b>	<b>Contenu de la thèse</b>	<b>14</b>
1.2.1	Aperçu général	14
1.2.2	Motivations et contributions	14

## 1.1 Cadre mathématique

D'abord, on pose les rudiments de l'estimation fonctionnelle. Ensuite, on présente les approches minimax et maxiset. Finalement, on précise les particularités de notre étude.

### 1.1.1 Point de départ

Cette sous-section rappelle les notions élémentaires de modèle statistique, d'estimateur et de risque.

#### 1.1.1.1 Modèle statistique

À la base de notre recherche, il y a la modélisation statistique de données observées. Cette modélisation, de nature plus ou moins complexe, est caractérisée par deux éléments. Le premier est une fonction inconnue  $f$ . Le second est une quantité d'information représentée par un entier  $n$ . Généralement, celui-ci précise le nombre de



données ou intervient sous la forme  $n^{-1}$  dans le niveau d'un bruit polluant l'émission d'une fonction.

Les modèles les plus étudiés en estimation fonctionnelle sont décrits ci-dessous.

**MODÈLE (DENSITÉS)**

*On observe  $n$  variables aléatoires i.i.d*

$$(X_1, \dots, X_n), \quad (1.1)$$

*où chaque  $X_i$  est de densité inconnue  $f$ .*

**MODÈLE (RÉGRESSION À PAS ÉQUIDISTANTS)**

*On observe  $n$  variables aléatoires i.i.d  $(Y_1, \dots, Y_n)$  avec*

$$Y_i = f(i/n) + z_i, \quad i = 1, \dots, n, \quad (1.2)$$

*où  $f : [0, 1] \rightarrow \mathbb{R}$  est une fonction inconnue bornée supérieurement et  $(z_1, \dots, z_n)$  sont des variables aléatoires i.i.d de loi  $\mathcal{N}(0, 1)$ .*

**MODÈLE (BRUIT BLANC GAUSSIEN)**

*On observe le processus continu  $\{Y(t); t \in [0, 1]\}$  défini par l'équation stochastique*

$$dY(t) = f(t)dt + n^{-1/2}dW(t), \quad (1.3)$$

*où  $f : [0, 1] \rightarrow \mathbb{R}$  est une fonction inconnue appartenant à  $\mathbb{L}^2([0, 1])$  et  $\{W(t); t \in [0, 1]\}$  est un mouvement Brownien standard.*

L'objectif est de reconstruire le plus fidèlement possible  $f$  à partir des observations. L'entier  $n$  est aussi grand qu'on le souhaite.

Dorénavant, on suppose que  $f$  appartient à  $\mathbb{L}^p([0, 1])$  avec  $p \in [1, \infty[$ , c'est-à-dire,  $\|f\|_p^p = \int_0^1 |f(x)|^p dx < \infty$ .

### 1.1.1.2 Estimateur

**DÉFINITION (ESTIMATEUR)**

*Toute fonction des observations est appelée estimateur.*

Pour juger de la précision d'un estimateur, des repères théoriques fiables doivent être mis en place. Telles sont les perspectives des approches minimax et maxiset. Pour les mettre en oeuvre, plusieurs objets mathématiques sont à choisir. Parmi ceux-ci, il y a le risque.

### 1.1.1.3 Risque

Le rôle d'un risque est de mesurer l'erreur commise par un estimateur dans la reconstruction de  $f$ . Dans notre étude, nous travaillons exclusivement avec le risque  $\mathbb{L}^p$ .

DÉFINITION (RISQUE  $\mathbb{L}^p$ )

Pour tout estimateur  $\hat{f}_n : [0, 1] \rightarrow \mathbb{R}$  de  $f$ , le risque  $\mathbb{L}^p$  de  $\hat{f}_n$  est défini par

$$\mathcal{L}_p(\hat{f}_n, f) = \mathbb{E}_f^n(\|\hat{f}_n - f\|_p^p),$$

où  $\mathbb{E}_f^n$  désigne l'espérance par rapport à la loi de probabilité de nos observations.

Les valeurs du paramètre  $p$  serviront à mettre en relief les propriétés locales de l'estimateur considéré. Par souci d'uniformité, tout ce qui suit sera présenté avec ce risque.

## 1.1.2 Approche minimax

L'approche minimax est un pilier de l'estimation fonctionnelle. Ses fondements ont été posés par Wolfowitz (1950). Une théorie solide a pris forme grâce aux travaux de Ibragimov et Hasminskii (1977), Speckman (1979), Bretagnolle et Carol-Huber (1979), Pinsker (1980), Sacks et Ylvisaker (1981), Efroimovich et Pinsker (1981), Stone (1982) et Birgé (1983).

### 1.1.2.1 Vitesses minimax

Pour commencer, on se fixe un ensemble de fonctions  $\mathcal{A}^\alpha \subset \mathbb{L}^p([0, 1])$ . Le paramètre  $\alpha$  caractérise la régularité des fonctions considérées. On suppose que la fonction inconnue appartient à cet ensemble. Enfin, on étudie le comportement asymptotique du risque minimax.

DÉFINITION (RISQUE MINIMAX)

Soit  $\mathcal{A}^\alpha$  un ensemble de fonctions. On définit le risque minimax par

$$\mathcal{R}_{n,p}(\mathcal{A}^\alpha) = \inf_{\hat{f}_n} \sup_{f \in \mathcal{A}^\alpha} \mathbb{E}_f^n(\|\hat{f}_n - f\|_p^p),$$

où l'infimum est pris sur l'ensemble des estimateurs  $\hat{f}_n$  de  $f$ .

Ainsi, on s'intéresse à la convergence du risque  $\mathbb{L}^p$  associé au meilleur estimateur pour la pire des fonctions appartenant à  $\mathcal{A}^\alpha$ . Trois définitions liées à cette convergence sont présentées ci-dessous.

DÉFINITION (BORNE INFÉRIEURE)

On appelle borne inférieure toute suite  $v_n$  positive telle que, pour  $n$  suffisamment grand, il existe une constante  $c > 0$  vérifiant

$$\mathcal{R}_{n,p}(\mathcal{A}^\alpha) \geq cv_n.$$

**DÉFINITION (BORNE SUPÉRIEURE)**

On appelle borne supérieure toute suite  $V_n$  positive telle que, pour  $n$  suffisamment grand, il existe une constante  $C > 0$  vérifiant

$$\mathcal{R}_{n,p}(\mathcal{A}^\alpha) \leqslant CV_n.$$

**DÉFINITION (VITESSE MINIMAX)**

Si il existe deux constantes  $C > 0$  et  $c > 0$  telles que, pour  $n$  suffisamment grand, une suite positive  $V_n$  vérifie

$$cV_n \leqslant \mathcal{R}_{n,p}(\mathcal{A}^\alpha) \leqslant CV_n,$$

alors  $V_n$  est appelée vitesse minimax.

La contrainte  $f \in \mathcal{A}^\alpha$  est primordiale. En effet, si aucune hypothèse de régularité n'est faite sur  $f$ , alors on ne peut généralement pas obtenir des résultats de convergence pour la suite  $\mathcal{R}_{n,p}(\mathbb{L}^p([0, 1]))$ . Cela a été montré par Farrell (1967).

Les principaux ensembles de fonctions rencontrés dans la littérature sont les boules de Sobolev, les boules de Hölder et, plus généralement, les boules de Besov.

Les vitesses minimax sont des repères théoriques servant à étudier les performances des estimateurs de  $f$ . Elles ont été calculées pour de nombreux modèles, risques et ensembles de fonctions. Par exemple, si on considère le risque  $\mathbb{L}^2$  et une boule de Sobolev de régularité  $\beta > 0$  alors, pour chacun des trois modèles (1.1), (1.2) et (1.3), la vitesse minimax est  $V_n = n^{-2\beta/(2\beta+1)}$ . Voir le livre de Tsybakov (2004).

### 1.1.2.2 Estimation adaptative

**DÉFINITION (ESTIMATEUR OPTIMAL)**

Soit  $V_n$  la vitesse minimax. Si il existe une constante  $C > 0$  telle que, pour  $n$  suffisamment grand, un estimateur  $\hat{f}_n$  vérifie

$$\sup_{f \in \mathcal{A}^\alpha} \mathbb{E}_f^n(\|\hat{f}_n - f\|_p^p) \leqslant CV_n,$$

alors il est dit optimal.

Pour déterminer une vitesse minimax, on est amené à construire un estimateur optimal. Généralement, ce dernier dépend explicitement du paramètre de régularité  $\alpha$ . Or la connaissance a priori de cette information est incompatible avec le fait que  $f$  soit inconnue. C'est pourquoi on cherche à élaborer des estimateurs optimaux complètement déterminés par les observations et aucunement de  $\alpha$ . De tels estimateurs sont dits adaptatifs.

Les travaux pionniers traitant de l'adaptation sont ceux de Efroimovich et Pinsker (1984), Golubev (1987) et Lepskii (1990; 1991). Depuis, de nombreuses méthodes

performantes ont fait leurs apparitions. Il y a notamment les méthodes de seuillage en ondelettes introduites par Donoho et Johnstone (1994; 1995), Donoho et al. (1995; 1996), les méthodes de noyaux à fenêtre adaptative élaborées par Lepskii et al. (1997) et la sélection de modèles par minimum de contraste pénalisé mise en place par Birgé et Massart (1997).

**BILAN DE L'APPROCHE MINIMAX :** On se fixe un ensemble de fonctions  $\mathcal{A}^\alpha$ . Les objectifs sont de calculer la vitesse minimax  $V_n$  et de construire des estimateurs adaptatifs optimaux.

L'approche minimax suscite quelques interrogations. Comment peut-on départager plusieurs estimateurs optimaux sur un même ensemble de fonctions  $\mathcal{A}^\alpha$  ? Est-ce qu'un estimateur optimal sur  $\mathcal{A}^\alpha$  l'est encore sur un ensemble de fonctions plus large ? Ou encore : quel est l'ensemble des fonctions qui seront bien estimées ou mal estimées par l'estimateur considéré ?

L'approche maxiset a pour optique de répondre à ces questions. Son principe vise à s'affranchir de la subjectivité de l'ensemble  $\mathcal{A}^\alpha$ .

### 1.1.3 Approche maxiset

L'approche maxiset a pris forme dans l'article de Cohen et al. (2000b). Depuis, elle a connu un développement significatif grâce aux articles de Kerkycharian et Picard (2000; 2002), la thèse de Rivoirard (2004) et la thèse de Autin (2005).

#### 1.1.3.1 Notion de maxiset

Pour commencer, on se fixe une suite positive  $w_n$  et un estimateur  $\hat{f}_n$  de  $f$ . Puis, on cherche à isoler l'ensemble des fonctions de taille maximale auquel appartient  $f$  pour que la borne supérieure de  $\hat{f}_n$  soit  $w_n$ .

**DÉFINITION (MAXISET)**

*Soit  $w_n$  une suite positive. On appelle maxiset d'un estimateur  $\hat{f}_n$  l'ensemble des fonctions  $\mathcal{B} \subseteq \mathbb{L}^p([0, 1])$  tel que, pour  $n$  suffisamment grand, il existe une constante  $C > 0$  vérifiant l'équivalence*

$$\mathbb{E}_f^n(\|\hat{f}_n - f\|_p^p) \leq Cw_n \iff f \in \mathcal{B}.$$

*Dorénavant, un tel maxiset est noté  $\mathcal{M}(\hat{f}_n, p, w_n)$ .*

Le facteur  $p$  signifie que l'on travaille avec le risque  $\mathbb{L}^p$ . L'expression finale d'un maxiset ne dépend pas de l'entier  $n$ . La présence de  $w_n$  dans la notation  $\mathcal{M}(\hat{f}_n, p, w_n)$  est informative.

En général, on s'intéresse aux maxiset associés aux estimateurs adaptatifs ayant déjà fait leurs preuves dans certains contextes minimax.

Si on prend  $w_n$  égale à la vitesse minimax sur un ensemble de fonctions  $\mathcal{A}^\alpha$  alors il existe deux constantes  $C > 0$  et  $c > 0$  telles que, pour  $n$  suffisamment grand, on a

$$cw_n \leq \mathcal{R}_{n,p}(\mathcal{M}(\hat{f}_n, p, w_n)) \leq Cw_n.$$

On peut donc mettre en relief l'optimalité d'un estimateur grâce à l'approche maxiset.

**ISOLEMENT D'UN MAXISET :** Soient  $\hat{f}_n$  un estimateur de  $f$  et  $w_n$  une suite positive. Pour déterminer  $\mathcal{M}(\hat{f}_n, p, w_n)$ , on cherche à identifier des quantités  $\aleph_n(f)$  et  $\aleph_n^*(f)$  dépendantes de  $f$  et  $n$  telles que, pour  $n$  suffisamment grand, il existe une constante  $C > 0$  vérifiant

$$\begin{cases} \aleph_n(f) \leq C(\mathbb{E}_f^n(\|\hat{f}_n - f\|_p^p) + w_n), \\ \mathbb{E}_f^n(\|\hat{f}_n - f\|_p^p) \leq C(\aleph_n^*(f) + w_n) \end{cases}$$

et

$$\aleph_n(f) \leq Cw_n \implies \aleph_n^*(f) \leq Cw_n.$$

Dans ces conditions, le maxiset  $\mathcal{M}(\hat{f}_n, p, w_n)$  est l'ensemble des fonctions  $f \in \mathbb{L}^p([0, 1])$  vérifiant

$$\limsup_{n \rightarrow \infty} w_n^{-1} \aleph_n(f) < \infty.$$

### 1.1.3.2 Comparaison maxiset

Le but premier d'un maxiset est de déterminer l'ensemble exact de fonctions sur lequel un estimateur est performant à une vitesse de convergence donnée. Naturellement, plus un maxiset est large, plus la fonction inconnue a des chances d'être estimée à la vitesse voulue par l'estimateur considéré. Il en découle la définition suivante :

**DÉFINITION (MEILLEUR AU SENS MAXISET)**

*Partant d'une suite positive  $w_n$ , un estimateur  $\hat{f}_n$  est dit meilleur au sens maxiset qu'un autre estimateur  $\hat{f}_n^*$  si leurs maxisets respectifs vérifient*

$$\mathcal{M}(\hat{f}_n^*, p, w_n) \subseteq \mathcal{M}(\hat{f}_n, p, w_n).$$

*On dit que  $\hat{f}_n$  est strictement meilleur au sens maxiset que  $\hat{f}_n^*$  si l'inclusion précédente est stricte.*

Il n'est pas nécessaire d'avoir la forme exacte des maxisets pour pouvoir les comparer. En effet, si il existe deux ensembles de fonctions  $A$  et  $B$  tels que

$$\mathcal{M}(\hat{f}_n^*, p, w_n) \subseteq B \subseteq A \subseteq \mathcal{M}(\hat{f}_n, p, w_n),$$

## 1.1. CADRE MATHÉMATIQUE

---

alors  $\hat{f}_n$  est meilleur au sens maxiset que  $\hat{f}_n^*$  pour la vitesse  $w_n$ .

**BILAN DE L'APPROCHE MAXISET :** On se fixe une suite positive  $w_n$ . L'objectif est de construire des estimateurs adaptatifs ayant le plus gros maxiset  $\mathcal{M}(\hat{f}_n, p, w_n)$  possible.

TAB. 1.1 – Récapitulatif : enjeux des approches minimax et maxiset.

	On pose	On cherche
Minimax	Un ensemble de fonctions $\mathcal{A}^\alpha$ .	La vitesse minimax $V_n$ , Des estimateurs adaptatifs $\hat{f}_n$ tels que $\sup_{f \in \mathcal{A}^\alpha} \mathbb{E}_f^n(\ \hat{f}_n - f\ _p^p) \leqslant CV_n$ .
Maxiset	Une suite positive $w_n$ .	Des estimateurs adaptatifs $\hat{f}_n$ tels que $\mathcal{M}(\hat{f}_n, p, w_n)$ soit le plus large possible.

### 1.1.4 Particularités de notre étude

Notre étude se singularise sur quatre points.

- *Risque* : comme mentionné précédemment, nous travaillons avec le risque  $\mathbb{L}^p$  et pas seulement avec le risque  $\mathbb{L}^2$ .
- *Modèles statistiques* : nous considérons principalement le modèle de régression à pas aléatoires, le modèle de bruit blanc gaussien généralisé et le modèle de convolution en bruit blanc gaussien. Les deux premiers sont des variantes complexes de (1.2) et (1.3). Le dernier est un classique des problèmes de déconvolution.
- *Ensemble  $\mathcal{A}^\alpha$*  : lorsque l'approche minimax est adoptée, nous travaillons sur des ensembles de fonctions appartenant aux espaces de Besov.
- *Méthodes de reconstruction* : tous les estimateurs que nous utilisons reposent sur l'analyse en ondelettes. Ceux-ci ont l'avantage d'être optimaux ou presque sous le risque  $\mathbb{L}^p$  sur des ensembles de fonctions larges. Plus de détails sont donnés dans le chapitre suivant.

## 1.2 Contenu de la thèse

Cette section donne une vue d'ensemble du travail effectué. Après un bref aperçu général, on décrit les motivations et ce que nous apportons de nouveau à l'estimation fonctionnelle.

### 1.2.1 Aperçu général

Cette thèse est composée de six chapitres. Les deux premiers composent la partie introductive. Le dernier est conclusif. Les autres s'orientent sur deux axes de recherches principaux.

- *Premier axe.* Chapitres 3 et 4. Étude de modèles statistiques complexes : influence d'une fonction parasite dans l'estimation de la fonction inconnue.
- *Deuxième axe.* Chapitre 5. Estimation adaptative : performances des estimateurs de seuillage par blocs sous le risque  $\mathbb{L}^p$ .

Le Chapitre 2 explique l'intérêt des ondelettes en estimation fonctionnelle. Le Chapitre 3 traite des vitesses minimax sur des boules de Besov. Le Chapitre 4 propose une étude minimax sur des boules de Besov pondérées. Le Chapitre 5 est consacré aux performances minimax et maxisets de plusieurs estimateurs adaptatifs reposant sur le seuillage par blocs. Le Chapitre 6 pose le bilan et les perspectives de nos recherches.

Les motivations et contributions qui alimentent nos travaux sont décrites ci-après.

### 1.2.2 Motivations et contributions

#### PARTIE I : ESTIMATION FONCTIONNELLE ET FONCTIONS PARASITES

Dans les Chapitres 3 et 4, nous considérons le modèle de bruit blanc gaussien généralisé et le modèle de régression à pas aléatoires. Ceux-ci sont des variantes complexes de (1.3) et (1.2). Nous les décrivons ci-dessous.

##### MODÈLE (BRUIT BLANC GAUSSIEN GÉNÉRALISÉ)

*On observe le processus continu  $\{Y(t); t \in [0, 1]\}$  défini par l'équation stochastique*

$$dY(t) = f(t)v(t)^{-1}dt + n^{-1/2}dW(t), \quad (1.4)$$

*où  $f : [0, 1] \rightarrow \mathbb{R}$  est une fonction inconnue bornée supérieurement (de borne connue),  $v : [0, 1] \rightarrow \mathbb{R}$  est une fonction connue telle que  $1/v \in \mathbb{L}^2([0, 1])$  et  $\{W(t); t \in [0, 1]\}$  est un mouvement Brownien standard.*

### MODÈLE (RÉGRESSION À PAS ALÉATOIRES)

*On observe  $n$  paires de variables aléatoires i.i.d  $((X_1, Y_1), \dots, (X_n, Y_n))$  gouvernées par l'équation*

$$Y_i = f(X_i) + z_i, \quad i = 1, \dots, n, \quad (1.5)$$

*où  $f : [0, 1] \rightarrow \mathbb{R}$  est une fonction inconnue bornée supérieurement (de borne connue),  $(X_1, \dots, X_n)$  sont des variables aléatoires i.i.d avec  $X_1 \in [0, 1]$  et  $(z_1, \dots, z_n)$  sont des variables aléatoires i.i.d de loi  $\mathcal{N}(0, 1)$  indépendantes des  $(X_1, \dots, X_n)$ . La fonction  $g : [0, 1] \rightarrow \mathbb{R}^+$  désigne la densité de  $X_1$ .*

Notre objectif est d'étudier l'influence de la fonction secondaire,  $v$  ou  $g$  selon le modèle, dans l'estimation de  $f$ . Pour ce faire, nous adoptons l'approche minimax sous le risque  $\mathbb{L}^p$  sur deux ensembles de fonctions différents. Dans le Chapitre 3, nous considérons les boules de Besov standards. Dans le Chapitre 4, nous choisissons les boules de Besov pondérées.

#### 1.2.2.1 Analyse du Chapitre 3

##### MOTIVATIONS PRINCIPALES DU CHAPITRE 3

Le modèle de régression à pas aléatoires (1.4) est à l'origine de notre travail. À l'aide de calculs élémentaires, on peut montrer que la nature de  $g$  détermine la répartition spatiale des observations. En particulier, si  $g$  s'annule en un point alors les observations présentes dans le voisinage de celui-ci sont peu nombreuses. Cela dresse un obstacle pour la reconstruction fidèle de  $f$ , surtout si cette fonction varie brutalement sur ce voisinage. Pour mettre en relief l'influence exacte de  $g$  dans l'estimation de  $f$ , le calcul des vitesses minimax s'impose.

Ce problème a été largement traité lorsque l'on se place sous le risque ponctuel sur des boules de Hölder. Voir, par exemple, la thèse de Gaïffas (2005). Ici, nous étudions l'expression de ces vitesses sous le risque  $\mathbb{L}^p$  sur des boules de Besov. L'intérêt de ce travail est d'expliquer

- pourquoi on rencontre des difficultés pour construire des estimateurs performants sous le risque  $\mathbb{L}^p$  sur des ensembles de fonctions standards quand la densité  $g$  s'annule en un point,
- et en particulier,
- pourquoi les articles étudiant l'estimation adaptative de  $f$  via l'approche minimax sous le risque  $\mathbb{L}^2$  sur des boules de Hölder ont du mal à se défaire de l'hypothèse "g minorée".



La considération du modèle de bruit blanc gaussien généralisé est venue par la suite. Techniquement parlant, l'influence de  $v$  dans l'expression des vitesses minimax est plus facile à cerner. C'est pourquoi nous le traitons avant le modèle de régression à pas aléatoires.

### CONTRIBUTIONS DU CHAPITRE 3

D'abord, nous présentons les outils mathématiques permettant de calculer les vitesses minimax sous le risque  $\mathbb{L}^p$  sur des boules de Besov pour un large panel de modèles statistiques. Les bornes supérieures sont déterminées grâce à deux estimateurs non adaptatifs en ondelettes. Le premier est l'estimateur par projection. Le second est un estimateur de seuillage dur développé par Delyon et Juditsky (1996). Pour isoler les bornes inférieures, nous considérons un théorème de minoration élaboré par Tsybakov (2004).

Puis, comme annoncé précédemment, nous étudions les modèles (1.4) et (1.5). Dans chacun des cas, nous utilisons les résultats de la première partie pour montrer l'influence de la "fonction parasite" dans l'expression des vitesses minimax. En guise de conclusion, nous commentons les limites de l'approche minimax considérée.

PRÉCISIONS : Afin de clarifier une des principales contributions du chapitre, décrivons brièvement les résultats obtenus pour le modèle de régression à pas aléatoires (1.5). Dorénavant, les boules de Besov sont représentées par  $B_{\pi,r}^s(L)$  où  $s$  est le paramètre de régularité,  $\pi$  et  $r$  sont des paramètres de norme et  $L$  désigne le rayon.

En premier lieu, on suppose que la fonction  $g$  est bornée inférieurement. Nous calculons les vitesses minimax sur  $B_{\pi,r}^s(L)$  pour toutes les valeurs de  $s$ ,  $\pi$  et  $r$ . Notons  $V_n$  ces vitesses. En second lieu, nous nous intéressons à l'ensemble des fonctions  $g$  telles que ces vitesses minimax sont conservées. Notons  $A$  cet ensemble. Nous montrons que si  $\pi \geq p$  alors on a :

$$\{g; \quad 1/g \in \mathbb{L}^{p_*-1}([0, 1])\} \subseteq A,$$

où  $p_* = \max(p, 2)$ . Il existe donc des fonctions  $g$  non minorées pour lesquelles certaines vitesses minimax obtenues dans le cas minoré ne sont pas altérées. Finalement, nous supposons que  $1/g \notin \mathbb{L}^{p_*-1}([0, 1])$  et nous soulignons la difficulté rencontrée pour déterminer les vitesses minimax à l'aide des méthodes en ondelettes standards.

TRANSITION VERS LE CHAPITRE 4 : Ainsi, lorsque qu'une fonction parasite intervient dans un modèle statistique, il est parfois difficile de calculer les vitesses minimax. Du moins, sous le risque  $\mathbb{L}^p$  sur des boules de Besov. Cette absence de repère est un obstacle pour estimer convenablement la fonction inconnue. Pour contourner celui-ci, une reconfiguration de notre approche minimax est envisageable. Le chapitre suivant étudie l'alternative proposée par les boules de Besov pondérées.

### 1.2.2.2 Analyse du Chapitre 4

#### MOTIVATIONS PRINCIPALES DU CHAPITRE 4

Les boules de Besov pondérées ont été introduites en estimation fonctionnelle par Kerkycharian et Picard (2005) dans le cadre du modèle de régression à pas aléatoires. Elles ont vu le jour grâce à l'étude maxiset d'un estimateur adaptatif construit sur une base d'ondelettes déformées. Si on considère ces boules et le risque  $\mathbb{L}^p$ , cet estimateur a la particularité d'atteindre une vitesse de convergence stable pour une large classe de densités  $g$ . Celle-ci est liée à une condition développée par Muckenhoupt (1972). Elle inclut notamment des densités vérifiant  $1/g \notin \mathbb{L}^{p^*-1}([0, 1])$ .

Deux interrogations se posent.

- Quelles sont les vitesses minimax associées ?
- Est-ce que les estimateurs construits sur des bases d'ondelettes déformées sont performants en pratique ?

Nous apportons des réponses précises à ces questions.

#### CONTRIBUTIONS DU CHAPITRE 4

Dans un premier temps, nous étudions les vitesses minimax sous le risque  $\mathbb{L}^p$  sur les boules de Besov pondérées. Notre objectif est de fournir des résultats applicables à de nombreux modèles statistiques mettant en jeu une fonction parasite. Pour amorcer notre travail, on suppose que celle-ci vérifie une condition de type Muckenhoupt.

D'abord, nous évaluons les bornes supérieures grâce à l'estimateur par projection en ondelettes déformées. Puis, nous établissons les bornes inférieures en combinant un théorème de minoration élaboré par Tsybakov (2004) avec une version pondérée du théorème de Varshamov-Gilbert. Enfin, nous appliquons ces résultats au modèle de bruit blanc gaussien généralisé (1.4) et au modèle de régression à pas aléatoires (1.5). Contrairement à leurs consoeurs, nous montrons que les boules de Besov pondérées fournissent des vitesses minimax stables pour une large classe de fonctions parasites. Par la même occasion, nous concluons que l'estimateur de seuillage en ondelettes déformées proposé par Kerkycharian et Picard (2005) est presque optimal pour de nombreux modèles statistiques.

Dans un deuxième temps, notre attention se focalise sur le modèle de régression à pas aléatoires (1.5). Nous présentons quelques simulations graphiques de deux estimateurs adaptatifs en ondelettes déformées introduits dans l'article de Kerkycharian et Picard (2005). L'un est explicitement construit avec la densité  $g$ . L'autre, plus réaliste, est entièrement déterminé par les observations sans connaissance a priori de  $g$ .

Dans les deux cas, nous constatons le bon rendu visuel de ces constructions. Les simulations sont mises en oeuvre grâce au logiciel Matlab 7 enrichi de la librairie Wavelab.

## PARTIE II : ESTIMATION ADAPTATIVE EN ONDELETTES

### 1.2.2.3 Analyse du Chapitre 5

#### MOTIVATIONS PRINCIPALES DU CHAPITRE 5

Le Chapitre 5 s'inscrit dans une perspective différente des précédents. À partir d'un modèle statistique général, on cherche à construire des estimateurs adaptatifs performants au sens minimax et maxiset sous le risque  $\mathbb{L}^p$ . Pour atteindre cet objectif, nous considérons le seuillage par blocs en ondelettes. Nous explorons principalement :

- l'estimateur de seuillage global construit par Kerkycharian et al. (1996),
- une version  $\mathbb{L}^p$  de l'estimateur BlockShrink développé par Cai (1996).

L'intérêt que nous portons à ces constructions est détaillé ci-après.

Dans la suite de cette section, les notions d'optimalité, de vitesse minimax et de supériorité maxiset utilisées sont sous-entendues "pour de nombreux modèles statistiques".

- Dans un premier temps, on s'intéresse au point de vue minimax. Voici les motivations et les trois grandes questions auxquelles nous répondons :
  - L'estimateur de seuillage global est optimal sous le risque  $\mathbb{L}^p$  sur  $B_{\pi,r}^s(L)$  avec  $\pi \geq p$  pour de nombreux modèles statistiques. Voir, par exemple, les articles de Kerkycharian et al. (1996) et Pensky et Vidakovic (1998). Cependant, quelles sont les performances de cet estimateur sous le risque  $\mathbb{L}^p$  sur  $B_{\pi,r}^s(L)$  avec  $p > \pi$  ?
  - L'estimateur BlockShrink est optimal sous le risque  $\mathbb{L}^2$  sur  $B_{\pi,r}^s(L)$  avec  $\pi \geq 2$  et presque optimal avec  $1 < \pi < 2$  pour de nombreux modèles statistiques. Voir, par exemple, les articles de Cai (1996; 1997; 1999; 2002b) et Chicken (2003b). Cependant, peut-on étendre ce résultat sous le risque  $\mathbb{L}^p$  avec  $p \neq 2$  ?
  - L'estimateur de seuillage dur est sous-optimal sous le risque  $\mathbb{L}^2$  sur  $B_{\pi,r}^s(L)$  avec  $\pi \geq 2$  pour le modèle de régression à pas équidistants (1.2). Cela a été montré par Cai (2002b). Cependant, qu'en est-il si on se place sous le risque  $\mathbb{L}^p$  avec  $p \neq 2$  et si on traite de modèles statistiques plus complexes ?
- Les estimateurs de seuillage par blocs ont été étudiés via l'approche maxiset par Cohen et al. (2000b) et Autin (2005).

Entre autre, Autin (2005) a montré que la version  $\mathbb{L}^p$  de l'estimateur BlockShrink est meilleure au sens maxiset que l'estimateur de seuillage dur introduit par Donoho et Johnstone (1995). Cela est obtenu sous le risque de Besov  $B_{p,p}^0$  avec  $p \geq 2$  pour le modèle de bruit blanc gaussien et les vitesses  $w_n \in \{n^{-\alpha p/2}, (\log n)^{\alpha p/2} n^{-\alpha p/2}\}$  avec  $\alpha \in ]0, 1[$ . Cependant :

- est-ce que l'estimateur BlockShrink considéré est strictement meilleur au sens maxiset que l'estimateur de seuillage dur ?
- est-ce que cela peut-être étendu sous le risque  $\mathbb{L}^p$  pour un large panel de modèles statistiques ?

Nous apportons des réponses à ces deux questions. Précisons que le risque  $\mathbb{L}^p$  est plus naturel à utiliser que le risque  $B_{p,p}^0$ . De plus, si  $p \geq 2$ , pour tout estimateur  $\hat{f}_n$  de  $f$ , on a

$$\mathbb{E}_f^n(\|\hat{f}_n - f\|_{B_{p,p}^0}^p) \leq \mathbb{E}_f^n(\|\hat{f}_n - f\|_p^p).$$

En ce sens, le risque  $\mathbb{L}^p$  est plus puissant que le risque  $B_{p,p}^0$ . En contrepartie, son utilisation fait appel à des outils mathématiques pointus.

Un des problèmes moteurs de nos applications concerne le modèle de régression à pas aléatoires (1.4). Quand la loi des pas est uniforme, Chicken (2003b) a montré l'optimalité de l'estimateur BlockShrink sous le risque  $\mathbb{L}^2$  sur des boules de Hölder. Nous étendons ce résultat en considérant les approches minimax et maxiset sous le risque  $\mathbb{L}^p$  quand la densité des pas est connue et bornée.

### CONTRIBUTIONS DU CHAPITRE 5

En premier lieu, nous encadrons le risque  $\mathbb{L}^p$  d'une large famille d'estimateurs de seuillage par blocs sans aucune hypothèse de régularité sur la fonction inconnue. Pour ce faire, nous utilisons certaines propriétés géométriques des bases d'ondelettes en norme  $\mathbb{L}^p$  et un découpage adéquat. Il en découle plusieurs résultats minimax et maxisets. Nous les décrivons ci-après.

D'abord, nous déterminons les bornes supérieures de ces estimateurs sur  $B_{\pi,r}^s(L)$  avec  $\pi \geq p$ . Les vitesses de convergence obtenues sont minimax pour de nombreux modèles statistiques. Nous complétons ce résultat en considérant la classe  $B_{\pi,r}^s(L)$  avec  $p > \pi$  et deux estimateurs de seuillage par blocs bien particuliers. Il s'agit de l'estimateur de seuillage global et une version  $\mathbb{L}^p$  de l'estimateur BlockShrink. Nous montrons la sous-optimalité du premier et l'optimalité quasi-généralisée du second. Une étude annexe concernant la sous-optimalité de l'estimateur de seuillage dur sous le risque  $\mathbb{L}^p$  sur  $B_{\pi,r}^s(L)$  avec  $\pi \geq p$  est faite. Grâce à elle, nous concluons à la supériorité minimax de l'estimateur BlockShrink considéré sur l'estimateur de seuillage

dur.

Ensuite, nous isolons le maxiset associé à l'estimateur de seuillage global pour la vitesse de convergence  $w_n = n^{-\alpha p/2}$  avec  $\alpha \in ]0, 1[$ . Puis, nous prouvons que l'estimateur BlockShrink considéré est strictement meilleur au sens maxiset que l'estimateur de seuillage dur pour les vitesses  $w_n \in \{n^{-\alpha p/2}, (\log n)^{\alpha p/2} n^{-\alpha p/2}\}$  avec  $\alpha \in ]0, 1[$ .

Finalement, nous mettons en relief la souplesse de nos résultats en les appliquant à deux modèles complexes : le modèle de régression à pas aléatoires (1.5) et le modèle de convolution en bruit blanc gaussien. Ce dernier appartient à la famille des problèmes de déconvolution.

#### PUBLICATIONS ET PRÉPUBLICATIONS<sup>1</sup>.

Les Chapitres 3 et 4 contiennent deux articles acceptés pour publication.

- Chesneau, C. "*A maxiset approach of a gaussian noise model.*" (2005). TEST.
- Chesneau, C. "*Regression in random design : a minimax study.*" (2007). Statistics and Probability Letters, 77, 40-53.

La partie simulation du Chapitre 4 est un court extrait d'un article fait en commun avec Thomas Willer<sup>2</sup> :

- Chesneau, C. and Willer, T. "*Regression in random design : a simulation study.*" (2005).

Le Chapitre 5 contient deux prépublications de l'université Paris VI.

- Chesneau, C. "*On wavelet block thresholding estimator : minimax and maxisets approaches under the  $\mathbb{L}^p$  risk.*" (2006a). (Soumis)
- Chesneau, C. "*Regression in random design and wavelet block thresholding estimator.*" (2006b). (Prochainement soumis)

Ces articles ont été réécrits dans le cadre de cette thèse. Quelques résultats annexes ont été rajoutés.

\* \* \*

---

<sup>1</sup>Elles sont disponibles en version anglaise sur le site : <http://www.chesneau-stat.com/>

<sup>2</sup>Université Paris VII. Contact e-mail : willer@math.jussieu.fr

# Chapitre 2

## Ondelettes et estimation fonctionnelle

Ce chapitre précise l'intérêt des ondelettes dans notre étude statistique.

### Sommaire

<b>2.1</b>	<b>Introduction</b>	<b>21</b>
<b>2.2</b>	<b>Bases d'ondelettes</b>	<b>22</b>
2.2.1	Construction	22
2.2.2	Décomposition	22
2.2.3	Propriétés géométriques	23
2.2.4	Boules de Besov	24
<b>2.3</b>	<b>Estimateurs en ondelettes</b>	<b>25</b>
2.3.1	Troncature	25
2.3.2	Seuillage en ondelettes	26
<b>2.4</b>	<b>Premiers résultats</b>	<b>27</b>
2.4.1	Préliminaires	28
2.4.2	Premier résultat minimax	29
2.4.3	Premier résultat maxiset	29
2.4.4	Applications concises	30

## 2.1 Introduction

D'abord, on traite des bases d'ondelettes. On décrit leur construction et plusieurs de leurs propriétés géométriques en norme  $\mathbb{L}^p$ . Ensuite, on introduit les boules de Besov et on fait état de leur caractérisation en ondelettes. Puis, on présente un bref état de l'art concernant les méthodes d'ondelettes en estimation fonctionnelle. Finalement, on détaille les performances minimax et maxisets de l'estimateur de seuillage dur développé par Donoho et Johnstone (1994).

## 2.2 Bases d'ondelettes

Analyser un signal revient à en extraire les informations qui le compose sous formes de valeurs numériques caractéristiques. Ces dernières doivent être relativement peu nombreuses pour être exploitables. Les bases d'ondelettes ont été élaborées dans cette perspective. En une quinzaine d'années, elles sont devenues des outils incontournables en traitement du signal. Pour une exploration complète du sujet, voir le livre de Mallat (2001).

### 2.2.1 Construction

Dans la présente étude, nous travaillons avec une base d'ondelettes sur l'intervalle unité. La nature de celle-ci changera suivant le modèle statistique traité. Pour fixer les idées, décrivons brièvement celle construite par Cohen et al. (1993).

**DÉFINITION (BASE D'ONDELETTES SUR L'INTERVALLE UNITÉ)**

*Soit  $N \in \mathbb{N}^*$ . Soient  $\phi$  une fonction d'échelle et  $\psi$  une ondelette mère  $N$ -régulière toutes deux à support compact. On pose :*

$$\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k), \quad \psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k).$$

*Par une modification appropriée aux bornes 0 et 1, on montre l'existence d'un entier  $\tau$  tel que, pour tout entier  $j_1 \geq \tau$ , la famille*

$$\zeta = \{\phi_{j_1,k}(x), k = 0, \dots, 2^{j_1} - 1; \psi_{j,k}(x), j = j_1, \dots, \infty, k = 0, \dots, 2^j - 1\}, \quad (2.1)$$

*forme une base orthonormée de  $\mathbb{L}^2([0, 1])$ .*

Pour les définitions de fonction d'échelle, d'ondelette mère et de  $N$ -régularité, voir le livre de Mallat (2001).

D'autres exemples de bases d'ondelettes sur l'intervalle unité seront présentés ultérieurement. Dorénavant, on adopte les notations utilisées pour définir la base  $\zeta$ . Les fonctions  $\phi$  et  $\psi$  ne seront pas nécessairement à support compact.

### 2.2.2 Décomposition

Toute fonction  $f \in \mathbb{L}^2([0, 1])$  se décompose sur une base d'ondelettes adaptée à l'intervalle unité comme

$$f(x) = \sum_{k=0}^{2^{j_1}-1} \alpha_{j_1,k} \phi_{j_1,k}(x) + \sum_{j=j_1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k}(x), \quad (2.2)$$

où

$$\alpha_{j,k} = \int_0^1 f(x) \phi_{j,k}(x) dx \quad \text{et} \quad \beta_{j,k} = \int_0^1 f(x) \psi_{j,k}(x) dx.$$

La première composante de la somme (2.2) est un terme d'approximation et la deuxième composante est un terme de détail.

Les bases d'ondelettes ont la faculté de concentrer les principales caractéristiques d'une fonction dans un nombre réduit de coefficients. Ces derniers sont identifiables par leur taille : plus ils sont "gros", plus l'information qu'ils véhiculent est significative. Les plus petits d'entre eux correspondent aux détails négligeables. On peut les éliminer sans altérer l'essentiel de la fonction.

### 2.2.3 Propriétés géométriques

Certaines bases d'ondelettes possèdent des propriétés géométriques fortes en norme  $\mathbb{L}^p$ . Il y a notamment la propriété d'inconditionnalité et la propriété de Temlyakov. La première a été établie par Meyer (1990). La seconde prend ses racines dans l'article de Temlyakov (1997).

#### 2.2.3.1 Propriétés géométriques fortes

PROPRIÉTÉ (INCONDITIONNALITÉ)

Soit  $p \in ]1, \infty[$ . Posons  $\psi_{\tau-1,j} = \phi_{\tau,k}$ . Pour toute suite  $u = (u_{j,k})_{j,k}$ , on a

$$\left\| \sum_{j=\tau-1}^{\infty} \sum_{k=0}^{2^j-1} u_{j,k} \psi_{j,k} \right\|_p^p \asymp \left\| \left( \sum_{j=\tau-1}^{\infty} \sum_{k=0}^{2^j-1} |u_{j,k} \psi_{j,k}|^2 \right)^{1/2} \right\|_p^p. \quad (2.3)$$

À titre de comparaison, précisons que la base de Fourier n'est pas inconditionnelle pour la norme  $\mathbb{L}^p$  avec  $p \neq 2$ .

PROPRIÉTÉ (TEMLYAKOV)

Soient  $p \in [2, \infty[$  et  $\sigma \in [0, \infty[$ . Posons  $\psi_{\tau-1,j} = \phi_{\tau,k}$ . Pour tout sous-ensemble  $A$  dans  $\{\tau-1, \dots, \infty\}$  et tout sous-ensemble  $\mathcal{C}$  dans  $\{0, \dots, 2^j-1\}$ , on a

$$\left\| \left( \sum_{j \in A} \sum_{k \in \mathcal{C}} |2^{\sigma j} \psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \asymp \sum_{j \in A} \sum_{k \in \mathcal{C}} 2^{\sigma j p} \|\psi_{j,k}\|_p^p \quad (2.4)$$

Pour plus de détails concernant l'encadrement (2.4), voir Johnstone et al. (2004, Theorem 2). Lorsque  $\sigma = 0$ , celle-ci a été mise en relief par Cohen et al. (2000a, Lemma 5.1). Ce dernier généralise un résultat prouvé par Temlyakov (1997).

#### 2.2.3.2 Propriétés standards

PROPRIÉTÉ (CONCENTRATION)

Soient  $v \in ]0, \infty[$  et  $h \in \{\phi, \psi\}$ . Il existe une constante  $C > 0$  telle que

$$\sum_{k=0}^{2^j-1} |h_{j,k}(x)|^v \leq C 2^{jv/2}, \quad x \in [0, 1]. \quad (2.5)$$



**PROPRIÉTÉ (ÉLÉMENTAIRE)**

Soient  $p \in [1, \infty[$ ,  $h \in \{\phi, \psi\}$  et  $\theta \in \{\alpha, \beta\}$ . Pour tout entier  $j \geq \tau$ , on a

$$\left\| \sum_{k=0}^{2^j-1} \theta_{j,k} h_{j,k} \right\|_p^p \asymp 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\theta_{j,k}|^p. \quad (2.6)$$

Précisons que la base  $\zeta$  décrite par (2.1) vérifie les quatre propriétés décrites ci-dessus.

## 2.2.4 Boules de Besov

Dans un premier temps, on définit les boules de Besov. Puis, on montre le lien qui existe entre celles-ci et les bases d'ondelettes.

### 2.2.4.1 Boules de Besov et module de continuité

**DÉFINITION (BOULES DE BESOV)**

Soient  $L \in ]0, \infty[$ ,  $s \in ]0, \infty[$ ,  $\pi \in [1, \infty]$  et  $u = \lfloor s \rfloor + 1$ . Pour toute fonction mesurable  $f : [0, 1] \rightarrow \mathbb{R}$ , on pose

$$\Delta_h(f)(x) = f(x+h) - f(x) = (\mathcal{T}_h - I)(f)(x),$$

où  $\mathcal{T}_h$  désigne l'opérateur de translation de pas  $h$ . On pose

$$\Delta_h^u(f)(x) = (\mathcal{T}_h - I)^u(f)(x) = \sum_{k=0}^u \binom{u}{k} (-1)^{u-k} f(x+kh).$$

On définit le  $u$ -ième module de continuité par

$$\rho^u(t, f, \pi) = \sup_{|h| \leq t} \left( \int_{J_{u,h}} |\Delta_h^u(f)(x)|^\pi dx \right)^{1/\pi},$$

où  $J_{u,h} = \{x \in [0, 1] \mid x+uh \in [0, 1]\}$ . Une fonction  $f$  appartient à la boule de Besov  $B_{\pi,r}^s(L)$  si et seulement si

$$\|f\|_\pi + \begin{cases} \left( \int_0^1 (t^{-s} \rho^u(t, f, \pi))^r t^{-1} dt \right)^{1/r} \leq L, & \text{si } r \in [1, \infty[, \\ \sup_{t \in ]0,1]} t^{-s} \rho^u(t, f, \pi) \leq L, & \text{si } r = \infty. \end{cases}$$

Le paramètre  $s$  représente la régularité des fonctions appartenant à la classe  $B_{\pi,r}^s(L)$ . Les paramètres  $\pi$  et  $r$  spécifient les normes dans lesquelles on mesure cette régularité. Pour plus de détails, voir l'article de DeVore et Popov (1988).

Classiquement,  $B_{\infty,\infty}^s(L)$  est la boule de Hölder et  $B_{2,2}^s(L)$  est la boule de Sobolev.

Pour tout  $a, b \in \mathbb{R}$ , on adopte les notations  $(a)_+ = \max(a, 0)$ ,  $a \vee b = \max(a, b)$  et  $a \wedge b = \min(a, b)$ .

**PROPRIÉTÉ (INCLUSION)**

Soient  $\pi \in [1, \infty]$ ,  $s \in ]1/\pi, \infty]$  et  $r \in [1, \infty]$ . Les boules de Besov vérifient l'inclusion

$$B_{\pi,r}^s(L) \subseteq B_{p,\infty}^{s-(1/\pi-1/p)_+}(L).$$

### 2.2.4.2 Ondelettes et boules de Besov

Il est possible de définir les boules de Besov à l'aide de certaines bases d'ondelettes sur l'intervalle unité. On a alors la caractérisation suivante :

PROPRIÉTÉ (CARACTÉRISATION EN ONDELETTES)

Soient  $\pi \in [1, \infty]$  et  $s \in ]1/\pi, N[$ . Une fonction  $f$  appartient à la boule de Besov  $B_{\pi,r}^s(L)$  si et seulement si les coefficients d'ondelettes associés vérifient

$$\left( \sum_{k=0}^{2^\tau-1} |\alpha_{\tau,k}|^\pi \right)^{1/\pi} + \begin{cases} (\sum_{j=\tau}^{\infty} (2^{j(s+1/2)} (2^{-j} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^\pi)^{1/\pi})^r)^{1/r} \leq L_*, & \text{si } r < \infty, \\ \sup_{j \geq \tau} 2^{j(s+1/2)} (2^{-j} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^\pi)^{1/\pi} \leq L_*, & \text{si } r = \infty, \end{cases}$$

où  $L_* > 0$  est une constante proportionnelle à  $L$ .

À titre d'exemple, la propriété précédente est vérifiée pour la base  $\zeta$  décrite par (2.1). Pour plus d'informations, voir le livre de Meyer (1990).

## 2.3 Estimateurs en ondelettes

Cette section rappelle quelques résultats mettant en jeu les bases d'ondelettes en estimation fonctionnelle.

Les travaux pionniers traitant de l'estimation fonctionnelle par méthodes d'ondelettes sont ceux de Doukhan et Leon (1990), Walter (1992) et Kerkycharian et Picard (1993). Il était alors question d'estimateurs linéaires non adaptatifs. Donoho et Johnstone (1994; 1995) et Donoho et al. (1996) ont élaboré les premières constructions adaptatives utilisant des techniques de seuillage en ondelettes.

### 2.3.1 Troncature

La construction d'un estimateur en ondelettes suit le schéma suivant. D'abord, on décompose la fonction inconnue  $f$  sur une base d'ondelettes pour avoir les coefficients  $\alpha_{j,k}$  et  $\beta_{j,k}$ . Puis, on estime ces coefficients inconnus via les observations. On note  $\hat{\alpha}_{j,k}$  et  $\hat{\beta}_{j,k}$  les estimateurs correspondants. Ensuite, on considère une certaine transformation de  $\hat{\beta}_{j,k}$  selon une règle préétablie. Enfin, on reconstruit le tout en utilisant la base d'ondelettes initiale. On obtient un estimateur  $\hat{f}_n : [0, 1] \rightarrow \mathbb{R}$  de la forme

$$\hat{f}_n(x) = \sum_{k=0}^{2^{j_1}-1} \hat{\alpha}_{j_1,k} \phi_{j_1,k}(x) + \sum_{j=j_1}^{\infty} \sum_{k=0}^{2^j-1} \mathcal{T}(\hat{\beta}_{j,k}) \psi_{j,k}(x),$$

où  $j_1$  est un entier quelconque supérieur à  $\tau$  et  $\mathcal{T}(\hat{\beta}_{j,k})$  représente une certaine transformation de  $\hat{\beta}_{j,k}$ .

Généralement, les hypothèses de régularité faites sur  $f$  entraînent l'existence d'un entier  $j_2$  à partir duquel les coefficients d'ondelettes  $\beta_{j,k}$  ne contiennent aucune information caractéristique. Si on les estime alors on injecte inutilement une dose d'imprécision dans la reconstruction finale. C'est pourquoi on considère une troncature de la forme

$$\mathcal{T}(\hat{\beta}_{j,k}) = \Upsilon(\hat{\beta}_{j,k})1_{\{j \in \{j_1, \dots, j_2\}\}},$$

où  $\Upsilon(\hat{\beta}_{j,k})$  désigne une certaine transformation de  $\hat{\beta}_{j,k}$ .

Si  $\Upsilon(\hat{\beta}_{j,k}) = \hat{\beta}_{j,k}$  alors  $\hat{f}_n$  appartient à la classe des estimateurs linéaires. Pour de nombreux modèles statistiques, leurs performances sont limitées quand on se place sur des ensembles de fonctions larges. Cela a été mis en relief par Donoho et al. (1996) via les boules de Besov. L'alternative proposée par les méthodes dites non linéaires consiste à opérer une sélection fine des estimateurs  $\hat{\beta}_{j,k}$ . Il est alors question de seuillage.

### 2.3.2 Seuillage en ondelettes

Dans la littérature statistique, plusieurs techniques de seuillage ont été développées. Il y a notamment le seuillage local et le seuillage par blocs. L'avantage du seuillage en ondelettes est de fournir des estimateurs à la fois adaptatifs et optimaux ou presque sous le risque  $\mathbb{L}^p$  sur des ensembles de fonctions larges. Cela est une conséquence de la sélection naturelle opérée par les bases d'ondelettes.

#### 2.3.2.1 Seuillage local

Le seuillage local est le plus utilisé. Son principe est le suivant : pour tout  $j \in \{j_1, \dots, j_2\}$  et  $k \in \{0, \dots, 2^j - 1\}$ , on considère une certaine transformation de  $\hat{\beta}_{j,k}$  si et seulement si  $|\hat{\beta}_{j,k}|$  dépasse un seuil  $\lambda$  préalablement fixé. À titre d'exemples, il y a le seuillage dur défini pour tout  $\lambda > 0$  par

$$\Upsilon(\hat{\beta}_{j,k}) = \hat{\beta}_{j,k}1_{\{|\hat{\beta}_{j,k}| \geq \lambda\}},$$

et le seuillage doux, tous deux développés par Donoho et Johnstone (1994; 1995). Dans la littérature, de nombreuses variantes ont été proposées. Entre autres, il y a le seuillage firm élaboré par Gao et Bruce (1996), le seuillage SCAD étudié par Fan (1997) et le seuillage non-negative garrote mis au point par Gao (1998),

#### 2.3.2.2 Choix du seuil

L'expression du seuil joue un rôle crucial dans les performances des estimateurs de seuillage. Or un choix convenable de celui-ci dépend de la nature du modèle statistique abordé. Pour une première approche, notons le sous la forme

$$\lambda = \mu \sqrt{(u_j/n)}, \tag{2.7}$$

où  $\mu$  est un réel positif et  $u = (u_j)_{j \in \{j_1, \dots, j_2\}}$  est une suite positive, aléatoire ou non, telle que  $u_j \in \{0, \dots, \log n\}$ .

Si  $u_j = \log n$  alors  $\lambda$  devient le seuil universel introduit par Donoho et Johnstone (1994). En considérant des variables aléatoires i.i.d  $(z_1, \dots, z_n)$  de loi  $\mathcal{N}(0, n^{-1})$ , celui-ci est caractérisé par la convergence

$$\lim_{n \rightarrow \infty} n^{\mu^2/2-1} \sqrt{\log n} \mathbb{P}(\max_{i \in \{1, \dots, n\}} |z_i| > \lambda) \asymp 1.$$

Les méthodes visant à sélectionner le meilleur seuil  $\lambda$  sont nombreuses. Celle d'où émane le seuil universel se nomme VisuShrink. Il y a aussi les méthodes RiskShrink et SureShrink développées par Donoho et Johnstone (1994; 1995), les méthodes de cross-validation étudiées par Nason (1996), Weyrich et Warhola (1998) et Jansen (2001), les méthodes reposant sur les tests d'hypothèses mises au point par Abramovich et Benjamini (1996), les méthodes dites "de Lepski" décrites par Juditsky (1997) et les méthodes bayésiennes proposées par Chipman et al. (1997) et Abramovich et al. (1998).

### 2.3.2.3 Seuillage par blocs

L'idée du seuillage par blocs a été introduite par Efroimovich (1985) dans le cadre de l'analyse de Fourier. Elle a été adaptée au contexte de l'analyse en ondelettes par Kerkycharian et al. (1996). Son principe est le suivant : pour chaque valeur de  $j$ , on sélectionne l'ensemble des estimateurs  $(\hat{\beta}_{j,k})_{k \in \{0, \dots, 2^j-1\}}$  si et seulement si une certaine transformation de ceux-ci dépasse un seuil préalablement fixé. Plus précisément, si on travaille sous le risque  $\mathbb{L}^p$ , pour tout  $\lambda > 0$ , on considère l'opérateur

$$\Upsilon(\hat{\beta}_{j,k}) = \hat{\beta}_{j,k} 1_{\{(2^{-j} \sum_{k=0}^{2^j-1} |\hat{\beta}_{j,k}|^p)^{1/p} \geq \lambda\}}.$$

Le seuil généralement utilisé est de la forme

$$\lambda = \mu n^{-1/2},$$

où  $\mu$  est un réel positif. La construction obtenue est appelée estimateur de seuillage global. Les premiers estimateurs de seuillage par blocs localisés ont été développés par Hall et al. (1998; 1999) et Cai (1996; 1997). Ceux-ci se montrent performants aussi bien en théorie qu'en pratique.

## 2.4 Premiers résultats

Cette section présente les performances minimax et maxisets sous le risque  $\mathbb{L}^p$  de l'estimateur de seuillage dur.

### 2.4.1 Préliminaires

Dans un premier temps, on définit l'estimateur de seuillage dur. Dans un deuxième temps, on pose les hypothèses qui seront utilisées.

On considère un modèle général à partir duquel on peut estimer les coefficients d'ondelettes  $\alpha_{j,k}$  et  $\beta_{j,k}$  de la fonction inconnue  $f : [0, 1] \rightarrow \mathbb{R}$ . Les estimateurs correspondants sont notés  $\hat{\alpha}_{j,k}$  et  $\hat{\beta}_{j,k}$ .

#### 2.4.1.1 Estimateur de seuillage dur

DÉFINITION (ESTIMATEUR DE SEUILLAGE DUR)

Soient  $\nu \in ]0, 1]$  et  $\mu \in ]0, \infty[$ . Soient  $j_1$  et  $j_2$  des entiers tels que

$$j_1 = \tau \quad \text{et} \quad 2^{j_2} \asymp (n/\log n)^\nu.$$

On définit l'estimateur de seuillage dur sous-entendu "à seuil universel"  $\hat{f}_n^h : [0, 1] \rightarrow \mathbb{R}$  par

$$\hat{f}_n^h(x) = \sum_{k=0}^{2^{j_1}-1} \hat{\alpha}_{j_1,k} \phi_{j_1,k}(x) + \sum_{j=j_1}^{j_2} \sum_{k=0}^{2^j-1} \hat{\beta}_{j,k} 1_{\{|\hat{\beta}_{j,k}| \geq \mu \sqrt{(\log n/n)}\}} \psi_{j,k}(x). \quad (2.8)$$

Toutes les paramètres présents sont choisis indépendamment de la régularité de  $f$ . L'estimateur  $\hat{f}_n^h$  est donc adaptatif.

#### 2.4.1.2 Hypothèses

Les deux hypothèses invoquées dans ce chapitre sont décrites ci-dessous. La première est une inégalité de moments et la deuxième est une inégalité de grande déviation.

HYPOTHÈSE (01)

Posons  $\hat{\beta}_{j_1-1,k} = \hat{\alpha}_{j_1,k}$ . Il existe une constante  $C > 0$  telle que, pour tout  $j \in \{j_1 - 1, \dots, j_2\}$ ,  $k \in \{0, \dots, 2^j - 1\}$  et  $n$  suffisamment grand, on a

$$\mathbb{E}_f^n(|\hat{\beta}_{j,k} - \beta_{j,k}|^{2p}) \leq Cn^{-p}.$$

HYPOTHÈSE (02)

Il existe deux constantes  $\mu_2$  et  $C > 0$  telles que, pour tout  $j \in \{j_1, \dots, j_2\}$ ,  $k \in \{0, \dots, 2^j - 1\}$  et  $n$  suffisamment grand, on a

$$\mathbb{P}_f^n \left( |\hat{\beta}_{j,k} - \beta_{j,k}| \geq 2^{-1} \mu_2 \sqrt{(\log n/n)} \right) \leq Cn^{-p}.$$

Le facteur  $p \in [1, \infty[$  correspond à celui du risque  $\mathbb{L}^p$  considéré. Ces deux hypothèses sont vérifiables pour de nombreux modèles statistiques. Quelques applications sont données à la fin de ce chapitre.

### 2.4.2 Premier résultat minimax

Les vitesses de convergence atteintes par l'estimateur de seuillage dur sous le risque  $\mathbb{L}^p$  sur des boules de Besov ont été synthétisées par Kerkyacharian et Picard (2000, Theorem 6.1).

**THÉORÈME 2.4.1 (KERKYACHARIAN ET PICARD (2000))**

*Soit  $p \in [1, \infty[$ . Plaçons nous dans le cadre statistique général décrit dans la sous-section 2.4.1. Supposons que les hypothèses (01) et (02) soient satisfaites. Considérons l'estimateur de seuillage dur  $\hat{f}_n^h$  défini par (3.2) avec la constante de seuillage  $\mu = \mu_2$ . Alors il existe une constante  $C > 0$  telle que, pour tout  $\pi \in [1, \infty]$ ,  $s \in ]1/\pi - (1/2 - 1/(2\nu))_+, N]$ ,  $r \in [1, \infty]$  et  $n$  suffisamment grand, on a*

$$\sup_{f \in B_{\pi, r}^s(L)} \mathbb{E}_f^n(\|\hat{f}_n^h - f\|_p^p) \leq C\varphi_n,$$

où

$$\varphi_n = \begin{cases} (\log n/n)^{\alpha_1 p}, & \text{si } \epsilon > 0, \\ (\log n/n)^{\alpha_2 p} (\log n)^{(p-\pi/r)+1_{\{\epsilon=0\}}}, & \text{si } \epsilon \leq 0, \end{cases}$$

avec  $\alpha_1 = s/(2s+1)$ ,  $\alpha_2 = (s - 1/\pi + 1/p)/(2(s - 1/\pi) + 1)$  et  $\epsilon = \pi s + 2^{-1}(\pi - p)$ .

L'estimateur de seuillage dur est presque optimal pour de nombreux modèles statistiques. La nuance "presque" est due à la présence du facteur logarithmique lorsque  $\epsilon > 0$ . Plus de précisions seront apportées dans le chapitre suivant.

**REMARQUE :** La première version du Théorème 2.4.1 a été montrée par Donoho et al. (1996, Theorem 3) dans le cadre du modèle de densités (1.1).

### 2.4.3 Premier résultat maxiset

Le maxiset associé à l'estimateur de seuillage dur sous le risque  $\mathbb{L}^p$  pour la vitesse  $w_n = (\log n/n)^{\alpha p/2}$  avec  $\alpha \in ]0, 1[$  a été déterminé par Kerkyacharian et Picard (2000, Theorems 5.1 et 5.2).

**THÉORÈME 2.4.2 (KERKYACHARIAN ET PICARD (2000))**

*Soit  $p \in ]1, \infty[$ . Plaçons nous dans le cadre statistique général décrit dans la sous-section 2.4.1. Supposons que les hypothèses (01) et (02) soient satisfaites. Considérons l'estimateur de seuillage dur  $\hat{f}_n^h$  défini par (3.2) avec la constante de seuillage  $\mu = \mu_2$ . Alors, pour tout  $\alpha \in ]0, 1[$ , on a*

$$\mathcal{M}(\hat{f}_n^h, p, (\log n/n)^{\alpha p/2}) = \mathcal{B}_{p, \infty}^{\alpha/(2\nu)} \cap \mathcal{W}((1-\alpha)p, p),$$

où

- l'espace  $\mathcal{B}_{p,\infty}^{\alpha/(2\nu)}$  est l'ensemble des fonctions  $f$  de  $\mathbb{L}^p([0, 1])$  tel que

$$\sup_{u>0} 2^{u\alpha p/(2\nu)} \left\| \sum_{j=u}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k} \right\|_p^p < \infty,$$

- l'espace  $\mathcal{W}((1-\alpha)p, p)$  est l'ensemble des fonctions  $f$  de  $\mathbb{L}^p([0, 1])$  tel que

$$\sup_{u>0} u^{-\alpha p} \left\| \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} 1_{\{|\beta_{j,k}| \leq u\}} \psi_{j,k} \right\|_p^p < \infty.$$

Le maxiset ainsi obtenu est large. Les inclusions suivantes ont été montrées par Kerkyacharian et Picard (2000, Theorem 6.1).

$$B_{\pi,r}^s(L) \subseteq \mathcal{B}_{p,\infty}^{\alpha_*/\nu} \cap \mathcal{W}((1-2\alpha_*)p, p) = \mathcal{M}(\hat{f}_n^h, p, (\log n/n)^{\alpha_* p}),$$

où  $\alpha_* = \alpha_1 1_{\{\epsilon > 0\}} + \alpha_2 1_{\{\epsilon \leq 0\}}$ ,  $\alpha_1 = s/(2s+1)$ ,  $\alpha_2 = (s-1/\pi+1/p)/(2(s-1/\pi)+1)$  et  $\epsilon = \pi s + 2^{-1}(\pi - p)$ . Par conséquent, on peut déduire les résultats minimax du Théorème 2.4.1 grâce au Théorème 2.4.2.

#### 2.4.4 Applications concises

Les Théorèmes 2.4.1 et 2.4.2 sont applicables à de nombreux modèles statistiques. En voici quelques exemples.

- Considérons le modèle de densités défini par (1.1). Les hypothèses (01) et (02) sont vérifiées pour  $\nu = 1$ ,

$$\hat{\alpha}_{j,k} = n^{-1} \sum_{i=1}^n \phi_{j,k}(X_i) \quad \text{et} \quad \hat{\beta}_{j,k} = n^{-1} \sum_{i=1}^n \psi_{j,k}(X_i).$$

Cela a été montré par Donoho et al. (1996).

- Considérons le modèle de régression à pas équidistants défini par (1.2). Les hypothèses (01) et (02) sont vérifiées pour  $\nu = 1$ ,

$$\hat{\alpha}_{j,k} = n^{-1} \sum_{i=1}^n Y_i \phi_{j,k}(i/n) \quad \text{et} \quad \hat{\beta}_{j,k} = n^{-1} \sum_{i=1}^n Y_i \psi_{j,k}(i/n).$$

Voir l'article de Donoho et Johnstone (1995).

- Considérons le modèle de bruit blanc gaussien défini par (1.3). Les hypothèses (01) et (02) sont vérifiées pour  $\nu = 1$ ,

$$\hat{\alpha}_{j,k} = \int_0^1 \phi_{j,k}(t) dY(t) \quad \text{et} \quad \hat{\beta}_{j,k} = \int_0^1 \psi_{j,k}(t) dY(t).$$

La preuve est standard.

## 2.4. PREMIERS RÉSULTATS

---

Ce qui précède constitue le point de départ de nos recherches. Les notions et notations introduites seront utilisées tout au long du manuscrit. Certains points tels le calcul des vitesses minimax ou le seuillage par blocs ont été abrégés car ils seront étudiés en détail dans les chapitres suivants.

\*            \*            \*





Chapitre

3

## Vitesses minimax : boules de Besov

Ce chapitre traite des vitesses minimax sous le risque  $\mathbb{L}^p$  sur des boules de Besov. L'objectif est d'étudier l'expression de ces vitesses pour certains modèles statistiques de nature complexe. Cette complexité est caractérisée par la présence d'une fonction secondaire pouvant perturber l'estimation de la fonction inconnue.

### Sommaire

<b>3.1</b>	<b>Introduction</b>	<b>33</b>
<b>3.2</b>	<b>Bornes supérieures et inférieures</b>	<b>34</b>
3.2.1	Bornes supérieures	34
3.2.2	Bornes inférieures	36
<b>3.3</b>	<b>Bruit blanc gaussien généralisé</b>	<b>37</b>
3.3.1	Présentation du modèle	37
3.3.2	Vitesses minimax	38
<b>3.4</b>	<b>Régression à pas aléatoires</b>	<b>41</b>
3.4.1	Présentation du modèle	41
3.4.2	Vitesses minimax	42
<b>3.5</b>	<b>Démonstrations</b>	<b>44</b>
3.5.1	Preuves des résultats de la Section 3.3	44
3.5.2	Preuves des résultats de la Section 3.4	52

### 3.1 Introduction

D'abord, nous présentons les outils mathématiques permettant de calculer les vitesses minimax sous le risque  $\mathbb{L}^p$  sur des boules de Besov pour un large panel de modèles statistiques. Les bornes supérieures sont déterminées grâce à deux estimateurs non adaptatifs en ondelettes. Le premier est l'estimateur par projection. Le second est un estimateur de seuillage dur développé par Delyon et Juditsky (1996). Pour isoler les bornes inférieures, nous utilisons un théorème de minoration élaboré par Tsybakov

(2004).

Puis, nous considérons le modèle du bruit blanc gaussien généralisé et le modèle de régression à pas aléatoires. Ceux-ci ont un point en commun : l'estimation de la fonction inconnue est perturbée par une fonction parasite. L'originalité de ce travail est de montrer l'influence de cette fonction dans l'expression des vitesses minimax. Nous montrons également les limites de l'approche minimax considérée. Cela servira de tremplin pour le chapitre suivant.

## 3.2 Bornes supérieures et inférieures

On considère un modèle général à partir duquel on peut estimer les coefficients d'ondelettes  $\alpha_{j,k}$  et  $\beta_{j,k}$  associés à la fonction inconnue  $f : [0, 1] \rightarrow \mathbb{R}$ . Les estimateurs correspondants sont notés  $\hat{\alpha}_{j,k}$  et  $\hat{\beta}_{j,k}$ .

### 3.2.1 Bornes supérieures

Cette sous-section traite des performances de l'estimateur par projection en ondelettes et de l'estimateur de seuillage dur développé par Delyon et Juditsky (1996). On considère l'approche minimax sous le risque  $\mathbb{L}^p$  sur des boules de Besov.

#### 3.2.1.1 Estimateur par projection

DÉFINITION (ESTIMATEUR PAR PROJECTION)

Soit  $j_1 \in \{\tau, \dots, \infty\}$ . On définit l'estimateur par projection en ondelettes  $\hat{f}_n^l : [0, 1] \rightarrow \mathbb{R}$  par

$$\hat{f}_n^l(x) = \sum_{k=0}^{2^{j_1}-1} \hat{\alpha}_{j_1,k} \phi_{j_1,k}(x). \quad (3.1)$$

Il a été introduit par Kerkycharian et Picard (1993).

Le Théorème 3.2.1 ci-dessous fournit une borne supérieure de l'estimateur linéaire sous le risque  $\mathbb{L}^p$  sur des boules de Besov.

THÉORÈME 3.2.1 (KERKYACHARIAN ET PICARD (1993))

Soit  $p \in [1, \infty[$ . Plaçons nous dans le cadre statistique général décrit au début de la présente section. Supposons que  $\hat{\alpha}_{j,k}$  et  $j_1$  vérifient

HYPOTHÈSE (A) Il existe deux constantes  $C > 0$  et  $\eta \geq 0$  telles que, pour  $n$  suffisamment grand, on a

$$\sum_{k=0}^{2^{j_1}-1} \mathbb{E}_f^n(|\hat{\alpha}_{j_1,k} - \alpha_{j_1,k}|^p) \leq C 2^{\eta j_1} n^{-p/2}.$$

### 3.2. BORNES SUPÉRIEURES ET INFÉRIEURES

---

Considérons l'estimateur  $\hat{f}_n^l$  défini par (3.1). Alors il existe une constante  $C > 0$  telle que, pour tout  $\pi \in [1, \infty]$ ,  $s \in ]1/\pi, N]$ ,  $r \in [1, \infty]$  et  $n$  suffisamment grand, on a

$$\sup_{f \in B_{\pi,r}^s(L)} \mathbb{E}_f^n(\|\hat{f}_n^l - f\|_p^p) \leq C\varphi_n,$$

où

$$\varphi_n = n^{-p/2} 2^{j_1(p/2-1+\eta)} + 2^{-j_1 s p} 2^{j_1(p-\pi)_+/\pi}.$$

L'atout du Théorème 3.2.1 réside dans sa souplesse. L'hypothèse (A) est large et l'entier  $j_1$  apparaissant dans la borne supérieure est complètement déterminé par l'utilisateur.

REMARQUE : Pour de nombreux modèles statistiques, l'estimateur par projection en ondelettes n'atteint pas les vitesses minimax sur  $B_{\pi,r}^s(L)$  pour toutes les valeurs de  $s$ ,  $\pi$  et  $r$ . Voir, par exemple, l'article de Donoho et al. (1996, Theorem 1).

Une alternative utilisant le seuillage en ondelettes est présentée dans la sous-section suivante.

#### 3.2.1.2 Estimateur de Delyon et Juditsky

DÉFINITION (ESTIMATEUR DE DELYON ET JUDITSKY)

Supposons que  $f \in B_{\pi,r}^s(L)$ . Soient  $\nu \in ]0, 1]$  et  $\mu \in ]0, \infty[$ . Soient  $j_1$ ,  $j_s$  et  $j_2$  des entiers tels que

$$j_1 = \tau, \quad 2^{j_s} \asymp n^{1/(2s+1)} \quad \text{et} \quad 2^{j_2} \asymp (n/\log n)^\nu.$$

On définit l'estimateur de Delyon et Juditsky  $\hat{f}_n^d : [0, 1] \rightarrow \mathbb{R}$  par

$$\hat{f}_n^d(x) = \sum_{k=0}^{2^{j_1}-1} \hat{\alpha}_{j_1,k} \phi_{j_1,k}(x) + \sum_{j=j_1}^{j_2} \sum_{k=0}^{2^j-1} \hat{\beta}_{j,k} 1_{\{|\hat{\beta}_{j,k}| \geq \mu \sqrt{((j-j_s)_+/n)}\}} \psi_{j,k}(x). \quad (3.2)$$

Comme il dépend du paramètre de régularité  $s$ , l'estimateur  $\hat{f}_n^d$  est non adaptatif.

Les hypothèses utilisées dans cette section sont décrites ci-dessous.

HYPOTHÈSE (B1)

Posons  $\hat{\beta}_{j_1-1,k} = \hat{\alpha}_{j_1,k}$ . Il existe une constante  $C > 0$  telle que, pour tout  $j \in \{j_1 - 1, \dots, j_2\}$ ,  $k \in \{0, \dots, 2^j - 1\}$  et  $n$  suffisamment grand, on a

$$\mathbb{E}_f^n(|\hat{\beta}_{j,k} - \beta_{j,k}|^{2p}) \leq Cn^{-p}.$$

HYPOTHÈSE (B2)

Il existe deux constantes  $C > 0$  et  $\mu_2 > 0$  telles que, pour tout  $a, j \in \{j_1 - 1, \dots, j_2\}$ ,  $k \in \{0, \dots, 2^j - 1\}$  et  $n$  suffisamment grand, on a

$$\mathbb{P}_f^n \left( |\hat{\beta}_{j,k} - \beta_{j,k}| \geq 2^{-1} \mu_2 \sqrt{(a/n)} \right) \leq C2^{-2ap}$$

Le facteur  $p \in [1, \infty[$  correspond à celui du risque  $\mathbb{L}^p$  considéré.

Le Théorème 3.2.2 ci-dessous calcule les vitesses de convergence atteintes par l'estimateur de Delyon et Juditsky sous le risque  $\mathbb{L}^p$  sur des boules de Besov.

**THÉORÈME 3.2.2** (DONOHO ET AL. (1996), DELYON ET JUDITSKY (1996))

*Soit  $p \in [1, \infty[$ . Plaçons nous dans le cadre statistique général décrit au début de la présente section. Supposons que les hypothèses (B1) et (B2) soient satisfaites. Considérons l'estimateur  $\hat{f}_n^d$  défini par (3.2) avec la constante de seuillage  $\mu = \mu_2$ . Alors il existe une constante  $C > 0$  telle que, pour tout  $\pi \in [1, \infty]$ ,  $s \in ]1/\pi - (1/2 - 1/(2\nu))_+, N]$ ,  $r \in [1, \infty]$  et  $n$  suffisamment grand, on a*

$$\sup_{f \in B_{\pi, r}^s(L)} \mathbb{E}_f^n(\|\hat{f}_n^d - f\|_p^p) \leq C\varphi_n,$$

où

$$\varphi_n = \begin{cases} n^{-\alpha_1 p}, & \text{si } \epsilon > 0, \\ (\log n/n)^{\alpha_2 p} (\log n)^{(p-\pi/r)+1_{\{\epsilon=0\}}}, & \text{si } \epsilon \leq 0, \end{cases}$$

avec  $\alpha_1 = s/(2s+1)$ ,  $\alpha_2 = (s-1/\pi+1/p)/(2(s-1/\pi)+1)$  et  $\epsilon = \pi s + 2^{-1}(\pi - p)$ .

Le Théorème 3.2.2 se montre en combinant deux résultats obtenus par Donoho et al. (1996, Theorem 3) et Delyon et Juditsky (1996, Theorem 1).

Les vitesses du Théorème 3.2.2 sont minimax pour de nombreux modèles statistiques. Des exemples précis sont traités dans les Sections 3.3 et 3.4.

L'optimalité de  $\hat{f}_n^d$  est due à la définition du seuil, lequel est moins rude que le seuil universel.

### 3.2.2 Bornes inférieures

Il existe de nombreuses techniques visant à minorer le risque minimax. Lorsque les risques  $\mathbb{L}^p$  et  $B_{\pi, r}^s(L)$  sont considérés, il est coutume d'appliquer deux résultats distincts suivant les valeurs de  $s$ ,  $\pi$  et  $r$ . Voir le livre de Härdle et al. (1998, Lemmas 10.1 et 10.2).

Le Théorème 3.2.3 ci-dessous est un cas particulier d'un résultat prouvé par Tsybakov (2004, Théorème 2.5). Celui-ci a l'avantage de fournir à lui seul toutes les bornes inférieures souhaitées pour de nombreux modèles statistiques. La quantité  $\mathbb{P}_f^n$  désigne la loi de probabilité de nos observations.

**THÉORÈME 3.2.3** (TSYBAKOV (2004))

*Soient  $p \in [1, \infty[$  et  $m \in \mathbb{N} - \{0, 1\}$ . Supposons qu'un ensemble de fonctions  $\mathcal{E}$  contienne les éléments  $h_0, \dots, h_m$  tels que :*

### 3.3. BRUIT BLANC GAUSSIEN GÉNÉRALISÉ

---

HYPOTHÈSE (C1) *Il existe un réel  $v > 0$  tel que*

$$\|h_i - h_l\|_p \geq 2v > 0,$$

*pour tout  $i \neq l$  et  $i, l \in \{0, \dots, m\}$ ,*

HYPOTHÈSE (C2)  $\mathbb{P}_{h_i}^n \ll \mathbb{P}_{h_l}^n$  avec  $i \neq l, i, l \in \{0, \dots, m\}$  et il existe une constante  $p_0 \in ]0, 8^{-1}[$  telle que

$$\inf_{l \in \{0, \dots, m\}} m^{-1} \sum_{i \neq l} \mathbb{E}_{h_i}^n (\log(\Lambda(\mathbb{P}_{h_i}^n, \mathbb{P}_{h_l}^n))) \leq p_0 \log(m),$$

où  $\Lambda(\mathbb{P}_{h_i}^n, \mathbb{P}_{h_l}^n) = d\mathbb{P}_{h_i}^n / d\mathbb{P}_{h_l}^n$  désigne le rapport de vraisemblance entre les lois induites par  $h_i$  et  $h_l$ . Alors

$$\mathcal{R}_{n,p}(\mathcal{E}) \geq c(m, p_0) v^p,$$

où  $c(m, p_0) = m^{1/2} (1 + m^{1/2})^{-1} (1 - 2p_0 - 2\sqrt{p_0 / \log m}) > 0$ .

Par la suite, le Théorème 3.2.3 sera appliqué avec des boules de type Besov. Les quantités  $h_0, \dots, h_m, m$  et  $v$  dépendront de  $n$ .

TRANSITION : Dorénavant, on focalise notre attention sur deux modèles statistiques précis. Le premier est une généralisation du modèle de bruit blanc gaussien et le second est le modèle de régression à pas aléatoires. Chacun d'entre eux fait intervenir une fonction parasite pouvant altérer l'estimation de la fonction inconnue. Nous étudions l'influence de cette fonction parasite dans l'expression des vitesses minimax.

## 3.3 Bruit blanc gaussien généralisé

D'abord, on présente le modèle de bruit blanc gaussien généralisé. Puis, on s'intéresse au calcul des vitesses minimax sous le risque  $\mathbb{L}^p$  sur des boules de Besov.

### 3.3.1 Présentation du modèle

#### 3.3.1.1 Description

MODÈLE (BRUIT BLANC GAUSSIEN GÉNÉRALISÉ)

*On observe le processus continu  $\{Y(t); t \in [0, 1]\}$  défini par l'équation stochastique*

$$dY(t) = f(t)v(t)^{-1}dt + n^{-1/2}dW(t), \quad (3.3)$$

où  $f : [0, 1] \rightarrow \mathbb{R}$  est une fonction inconnue bornée supérieurement (de borne connue),  $v : [0, 1] \rightarrow \mathbb{R}$  est une fonction connue telle que  $1/v \in \mathbb{L}^2([0, 1])$  et  $\{W(t); t \in [0, 1]\}$  est un mouvement Brownien standard sur  $[0, 1]$ .

L'objectif est de reconstruire  $f$  à partir des observations  $\{\int_0^1 h(t)dY(t), h \in \mathbb{L}^2([0, 1])\}$ . Nous étudions le comportement des vitesses minimax sous le risque  $\mathbb{L}^p$  sur des boules de Besov pour des classes de fonctions  $v$  aussi larges que possible.

### 3.3.1.2 Commentaires

- Si  $v = 1$  alors (3.3) devient le modèle de bruit blanc gaussien. Celui-ci a été étudié par de nombreux auteurs, à commencer par Ibragimov et Hasminskii (1977). En raison de sa simplicité mathématique, il idéalise le problème général de reconstruction des fonctions bruitées. Pour une étude minimax complète, voir le livre de Tsybakov (2004).
- Si  $v$  est une fonction non constante alors la nature du modèle (3.3) se complexifie. En effet, si  $v \notin \mathbb{L}^2([0, 1])$  alors le terme  $v(t)dW(t)$  et, a fortiori, le modèle  $v(t)dY(t)$  n'ont pas de sens. Un obstacle de taille se dresse si on souhaite estimer les coefficients d'ondelettes associés à  $f$ .

À notre connaissance, le modèle (3.3) n'a jamais été étudié sous une forme aussi générale. Toutefois, lorsque des hypothèses de majoration et minoration sont faites sur  $v$ , un certain nombre de résultats ont été établis. Le paragraphe ci-dessous fait état de l'un d'entre eux.

Considérons le modèle de régression hétéroscédastique à pas aléatoires.

MODÈLE (RÉGRESSION HÉTÉROSCÉDASTIQUE À PAS ALÉATOIRES)

On observe  $n$  paires de variables aléatoires *i.i.d*  $((X_1, Y_1), \dots, (X_n, Y_n))$  avec

$$Y_i = f(X_i) + \sigma(X_i)z_i, \quad i = 1, \dots, n, \quad (3.4)$$

où  $f : [0, 1] \rightarrow \mathbb{R}$  est une fonction inconnue bornée supérieurement (de borne connue),  $(X_1, \dots, X_n)$  sont des variables aléatoires *i.i.d* de densité  $g : [0, 1] \rightarrow \mathbb{R}^+$ ,  $(z_1, \dots, z_n)$  sont des variables aléatoires *i.i.d* de loi  $\mathcal{N}(0, 1)$  et sont indépendantes des  $(X_1, \dots, X_n)$ .

Brown et Low (1996) ont montré que les modèles (3.4) et

$$v(t)dY(t) = f(t)dt + n^{-1/2}v(t)dW(t),$$

sont équivalents au sens de Le Cam à condition que  $v(t) = \sigma(t)g(t)^{-1/2}$ , que  $g$  et  $\sigma$  vérifient certaines hypothèses de bornitude et que  $f$  vérifie certaines hypothèses de régularité. Cela a trouvé une application dans l'article de Efroimovich et Pinsker (1996). Pour plus d'informations concernant l'équivalence de modèle au sens de Le Cam, voir le livre de Le Cam et Yang (1990).

### 3.3.2 Vitesses minimax

Considérons le modèle de bruit blanc gaussien généralisé défini par (3.3). En premier lieu, nous nous intéressons au cas où la fonction  $v$  est bornée supérieurement. Puis nous traitons le cas non bornée.

#### 3.3.2.1 Cas borné

Sous une hypothèse de bornitude sur  $v$ , le Théorème 3.3.1 ci-dessous établit les vitesses minimax sous le risque  $\mathbb{L}^p$  sur des boules de Besov.

##### THÉORÈME 3.3.1

*Considérons le modèle de bruit blanc gaussien généralisé défini par (3.3). Supposons que  $v$  soit bornée supérieurement. Soient  $p \in [1, \infty[$ ,  $\pi \in [1, \infty]$ ,  $s \in ]1/\pi, N]$  et  $r \in [1, \infty]$ . Posons  $\alpha_1 = s/(2s + 1)$ ,  $\alpha_2 = (s - 1/\pi + 1/p)/(2(s - 1/\pi) + 1)$  et  $\epsilon = \pi s + 2^{-1}(\pi - p)$ .*

- *Si  $\epsilon \neq 0$  alors il existe deux constantes  $C > 0$  et  $c > 0$  telles que, pour  $n$  suffisamment grand, on a*

$$c\varphi_n \leq \mathcal{R}_{n,p}(B_{\pi,r}^s(L)) \leq C\varphi_n,$$

où

$$\varphi_n = \begin{cases} n^{-\alpha_1 p}, & \text{si } \epsilon > 0, \\ (\log n/n)^{\alpha_2 p}, & \text{si } \epsilon < 0. \end{cases}$$

- *Si  $\epsilon = 0$  alors il existe deux constantes  $C > 0$  et  $c > 0$  telles que, pour  $n$  suffisamment grand, on a*

$$c\varphi_{n,1} \leq \mathcal{R}_{n,p}(B_{\pi,r}^s(L)) \leq C\varphi_{n,2},$$

où

$$\varphi_{n,i} = (\log n/n)^{\alpha_2 p} (\log n)^{(p-\pi/r)_+ 1_{\{i=2\}}}, \quad i \in \{1, 2\}.$$

Les bornes supérieures sont calculées grâce au Théorème 3.2.2. Les bornes inférieures découlent du Théorème 3.2.3. Ces dernières sont obtenues sans aucune condition sur  $v$  autre que  $1/v \in \mathbb{L}^2([0, 1])$ .

REMARQUE : Lorsque  $v = 1$ , on retrouve les vitesses minimax liées au modèle de bruit blanc gaussien standard. Voir, par exemple, l'article de Cai et al. (2005, Theorem 2).

Après analyse du Théorème 3.3.1, il est naturel de se poser la question suivante : comment se comportent les vitesses minimax sous le risque  $\mathbb{L}^p$  sur des boules de Besov si  $v$  n'est pas bornée supérieurement ? Des éléments de réponses sont apportés dans la sous-section suivante.

#### 3.3.2.2 Cas dégénéré

Le Théorème 3.3.2 ci-dessous montre que la vitesse minimax exhibée au Théorème 3.3.1 lorsque  $\pi \geq p$  est conservée pour des fonctions  $v$  non nécessairement majorées.



**THÉORÈME 3.3.2**

Soit  $p \in [1, \infty[$ . Considérons le modèle de bruit blanc gaussien généralisé défini par (3.3). Supposons que

$$v \in \mathbb{L}^{p_*}([0, 1]), \quad p_* = p \vee 2. \quad (3.5)$$

Alors il existe deux constantes  $C > 0$  et  $c > 0$  telles que, pour tout  $\pi \in [p, \infty]$ ,  $s \in ]1/\pi, N]$ ,  $r \in [1, \infty]$  et  $n$  suffisamment grand, on a

$$cn^{-\alpha_1 p} \leq \mathcal{R}_{n,p}(B_{\pi,r}^s(L)) \leq Cn^{-\alpha_1 p},$$

où  $\alpha_1 = s/(2s + 1)$ .

Les bornes supérieures sont calculées grâce au Théorème 3.2.1. Les bornes inférieures découlent du Théorème 3.2.3.

Si l'hypothèse (3.5) n'est pas vérifiée alors  $v$  peut altérer cette vitesse minimax. Cela est mis en relief dans la Proposition 3.3.1 ci-dessous.

**PROPOSITION 3.3.1**

Soit  $p \in [1, \infty[$ . Considérons le modèle de bruit blanc gaussien généralisé défini par (3.3). Il existe des fonctions  $v \notin \mathbb{L}^{p_*}([0, 1])$  avec  $p_* = p \vee 2$  pour lesquelles les vitesses minimax du Théorème 3.3.1 ne peuvent être atteintes par aucun estimateur. Cela est vrai pour toutes les valeurs de  $s$ ,  $\pi$  et  $r$ .

La preuve de la Proposition 3.3.1 repose sur le calcul d'une borne inférieure. Plus précisément, nous considérons la fonction  $v(t) = t^{-\sigma/2}$  avec  $\sigma > 2p^{-1}$  et nous appliquons le Théorème 3.2.3 pour montrer l'existence d'une constante  $c > 0$  telle que :

$$\mathcal{R}_{n,p}(B_{\pi,r}^s(L)) \geq cn^{-\alpha_3 p},$$

où  $\alpha_3 = (s - 1/\pi + 1/p)/(2(s - 1/\pi) + 1 + \sigma)$ . Celle-ci est supérieure aux vitesses de convergences décrites dans le Théorème 3.3.1. En particulier, on constate que plus  $\sigma$  est grand, plus  $\varphi_n$  est lente.

Le Théorème 3.3.3 ci-dessous complète la Proposition 3.3.1. Nous calculons la vitesse minimax sous le risque  $\mathbb{L}^p$  sur  $B_{\pi,r}^s(L)$  avec  $p \geq \pi$  pour une fonction  $v$  n'appartenant pas à  $\mathbb{L}^{p_*}([0, 1])$ .

**THÉORÈME 3.3.3**

Soit  $p \in ]2, \infty[$ . Considérons le modèle de bruit blanc gaussien généralisé défini par (3.3) avec

$$v(t) = t^{-\sigma/2}, \quad 2p^{-1} < \sigma < 1.$$

Alors il existe deux constantes  $C > 0$  et  $c > 0$  telles que, pour tout  $\pi \in [1, p]$ ,  $s \in ]1/\pi, N]$ ,  $r \in [1, \infty]$  et  $n$  suffisamment grand, on a

$$cn^{-\alpha_3 p} \leq \mathcal{R}_{n,p}(B_{\pi,r}^s(L)) \leq Cn^{-\alpha_3 p},$$

où  $\alpha_3 = (s - 1/\pi + 1/p)/(2(s - 1/\pi) + 1 + \sigma)$ .

Les bornes supérieures sont calculées grâce au Théorème 3.2.1. Les bornes inférieures découlent du Théorème 3.2.3.

Ce résultat soulève quelques interrogations et notamment : quelles sont les vitesses minimax lorsque  $\sigma > 1$  ? En raison de plusieurs difficultés d'ordre technique, nous ne pouvons pas apporter une réponse définitive à cette question.

De manière générale, lorsque  $v \notin \mathbb{L}^{p*}([0, 1])$ , il est difficile d'isoler les vitesses minimax sous le risque  $\mathbb{L}^p$  sur  $B_{\pi,r}^s(L)$  pour tout  $s, \pi$  et  $r$ . Par conséquent, l'estimation adaptative de  $f$  devient délicate. Une alternative utilisant les boules de Besov pondérées est développée dans le Chapitre 4.

## 3.4 Régression à pas aléatoires

D'abord, on présente le modèle de régression à pas aléatoires. Puis, on s'intéresse au calcul des vitesses minimax sous le risque  $\mathbb{L}^p$  sur des boules de Besov.

### 3.4.1 Présentation du modèle

#### 3.4.1.1 Description

MODÈLE (RÉGRESSION À PAS ALÉATOIRES)

*On observe  $n$  paires de variables aléatoires i.i.d  $((X_1, Y_1), \dots, (X_n, Y_n))$  gouvernées par l'équation*

$$Y_i = f(X_i) + z_i, \quad i = 1, \dots, n, \quad (3.6)$$

*où  $f : [0, 1] \rightarrow \mathbb{R}$  est une fonction inconnue bornée supérieurement (de borne connue),  $(X_1, \dots, X_n)$  sont des variables aléatoires i.i.d avec  $X_1 \in [0, 1]$  et  $(z_1, \dots, z_n)$  sont des variables aléatoires i.i.d de loi  $\mathcal{N}(0, 1)$  indépendantes des  $(X_1, \dots, X_n)$ . Dans notre étude, la fonction  $g : [0, 1] \rightarrow \mathbb{R}^+$  désigne la densité de  $X_1$ . Dans la plupart du temps, on la supposera connue.*

L'objectif est de reconstruire  $f$  à partir des observations  $((Y_1, X_1), \dots, (Y_n, X_n))$ .

Nous étudions le comportement des vitesses minimax sous le risque  $\mathbb{L}^p$  sur des boules de Besov pour une large classe de fonctions  $g$ . Les résultats obtenus sont analogues à ceux trouvés dans le cas du modèle de bruit blanc gaussien généralisé.

#### 3.4.1.2 Commentaires

Contrairement au modèle de bruit blanc gaussien généralisé (3.3), le modèle de régression à pas aléatoires (3.6) trouve des applications pratiques. En contrepartie, l'estimation de  $f$  fait appel à des outils mathématiques plus pointus.

- Si  $g$  vérifie certaines hypothèses de bornitude ou suit une loi uniforme alors ce problème à été étudié par de nombreux auteurs sous divers risques sur des ensembles de fonctions variés. Voir, par exemple, les articles de Ibragimov et Hasminskii (1982), Stone (1982), Delyon et Juditsky (1996), Cai et Brown (1999), Pensky et Vidakovic (2001), Delouille et al. (2004) et l'ouvrage de Tsybakov (2004).
- Quelques articles traitent du cas où  $g$  s'annule en un point via l'approche minimax sous le risque ponctuel et uniforme sur des boules de Hölder. Pour un aperçu complet, voir la thèse de Gaïffas (2005).

À notre connaissance, aucun travail n'a été fait sous le risque  $\mathbb{L}^p$  sur des boules de Besov. Pourtant, cela pourrait expliquer les difficultés que l'on rencontre pour construire des estimateurs adaptatifs performants dans ce contexte statistique, d'où la motivation de notre travail.

### 3.4.2 Vitesses minimax

Considérons le modèle de régression à pas aléatoires défini par (3.6). Dans un premier temps, on s'intéresse au cas où  $g$  est minorée. Puis on étudie le cas où  $g$  peut s'annuler en un point.

#### 3.4.2.1 Cas borné

Sous une hypothèse de bornitude sur  $g$ , le Théorème 3.4.1 ci-dessous établit les vitesses minimax sous le risque  $\mathbb{L}^p$  sur des boules de Besov.

##### THÉORÈME 3.4.1

*Considérons le modèle de régression à pas aléatoires défini par (3.6). Supposons que  $g$  soit bornée inférieurement. Soient  $p \in [1, \infty[$ ,  $\pi \in [1, \infty]$ ,  $s \in [1/2 + 1/\pi, N]$  et  $r \in [1, \infty]$ . Posons  $\alpha_1 = s/(2s + 1)$ ,  $\alpha_2 = (s - 1/\pi + 1/p)/(2(s - 1/\pi) + 1)$  et  $\epsilon = \pi s + 2^{-1}(\pi - p)$ .*

- Si  $\epsilon \neq 0$  alors il existe deux constantes  $C > 0$  et  $c > 0$  telles que, pour  $n$  suffisamment grand, on a

$$c\varphi_n \leq \mathcal{R}_{n,p}(B_{\pi,r}^s(L)) \leq C\varphi_n,$$

où

$$\varphi_n = \begin{cases} n^{-\alpha_1 p}, & \text{si } \epsilon > 0, \\ (\log n/n)^{\alpha_2 p}, & \text{si } \epsilon < 0. \end{cases}$$

- Si  $\epsilon = 0$  alors il existe deux constantes  $C > 0$  et  $c > 0$  telles que, pour  $n$  suffisamment grand, on a

$$c\varphi_{n,1} \leq \mathcal{R}_{n,p}(B_{\pi,r}^s(L)) \leq C\varphi_{n,2},$$

où

$$\varphi_{n,i} = (\log n/n)^{\alpha_2 p} (\log n)^{(p-\pi/r)_+ 1_{\{i=2\}}}, \quad i \in \{1, 2\}.$$

Les bornes supérieures sont calculées grâce au Théorème 3.2.2. Pour appliquer celui-ci, nous proposons des estimateurs  $\hat{\alpha}_{j,k}$  et  $\hat{\beta}_{j,k}$  vérifiant les hypothèses (B1) et (B2). L'inégalité de Rosenthal, l'inégalité de Hoeffding et l'inégalité de Bernstein sont utilisées. Les bornes inférieures découlent du Théorème 3.2.3. Ces dernières sont obtenues sans aucune condition particulière sur  $g$ .

#### 3.4.2.2 Cas dégénéré

Le Théorème 3.4.2 ci-dessous montre l'existence de fonctions  $g$  non minorées telles que la vitesse minimax exhibée au Théorème 3.4.1 pour le cas  $\pi \geq p$  est conservée.

##### THÉORÈME 3.4.2

Soit  $p \in [1, \infty[$ . Considérons le modèle de régression à pas aléatoires défini par (3.6). Supposons que

$$1/g \in \mathbb{L}^{p_*-1}([0, 1]), \quad p_* = p \vee 2. \quad (3.7)$$

Alors il existe deux constantes  $C > 0$  et  $c > 0$  telles que, pour tout  $\pi \in [p, \infty]$ ,  $s \in ]1/\pi, N]$ ,  $r \in [1, \infty]$  et  $n$  suffisamment grand, on a

$$cn^{-\alpha_1 p} \leq \mathcal{R}_{n,p}(B_{\pi,r}^s(L)) \leq Cn^{-\alpha_1 p},$$

où  $\alpha_1 = s/(2s+1)$ .

Les bornes supérieures sont calculées grâce au Théorème 3.2.1. Les bornes inférieures découlent du Théorème 3.2.3.

Si l'hypothèse (3.7) n'est pas vérifiée, alors  $g$  peut altérer cette vitesse minimax. Cela est précisé dans la Proposition 3.4.1 ci-dessous.

##### PROPOSITION 3.4.1

Soit  $p \in [1, \infty[$ . Considérons le modèle de régression à pas aléatoires défini par (3.6). Il existe des fonctions  $g$  vérifiant  $1/g \notin \mathbb{L}^{p_*-1}([0, 1])$  avec  $p_* = p \vee 2$  pour lesquelles les vitesses minimax du Théorème 3.3.1 ne peuvent être atteintes par aucun estimateur. Cela est vrai pour toutes les valeurs de  $s$ ,  $\pi$  et  $r$ .

La preuve de la Proposition 3.4.1 repose sur une borne inférieure montrée via le Théorème 3.2.3. Pour être plus précis, nous considérons la densité  $g(t) = (\sigma + 1)t^\sigma$  avec  $\sigma > 2p^{-1}$  et nous montrons l'existence d'une constante  $c > 0$  telle que :

$$\mathcal{R}_{n,p}(B_{\pi,r}^s(L)) \geq cn^{-\alpha_3 p},$$

avec  $\alpha_3 = (s - 1/\pi + 1/p)/(2(s - 1/\pi) + 1 + \sigma)$ . Remarquons que plus  $\sigma$  est grand, plus la vitesse minimax est lente. Cela rejoint ce que l'on observe en pratique à savoir : plus la fonction  $g$  s'écrase sur l'axe des abscisses au voisinage d'un point, moins

on dispose d'information sur  $f$  dans ce voisinage et, a fortiori, plus l'estimation de  $f$  se complique.

De manière générale, si  $1/g \notin \mathbb{L}^2([0, 1])$  alors il est difficile d'isoler les vitesses minimax dans le cadre statistique posé ci-dessus. La construction d'estimateurs adaptatifs performants s'avère compliquée. Une alternative utilisant les boules de Besov pondérées est considérée dans le chapitre suivant.

## 3.5 Démonstrations

Les constantes  $c$  et  $C$  ne dépendent ni de  $f$ , ni de  $n$ . Elles peuvent prendre des valeurs différentes d'un terme à l'autre. On suppose que  $n$  est suffisamment grand. Dans toute cette section, on travaille exclusivement avec la base d'ondelettes  $\zeta$  décrite par (2.1).

### 3.5.1 Preuves des résultats de la Section 3.3

**Preuve du Théorème 3.3.1.** Étudions séparément les bornes inférieures et les bornes supérieures.

- *Bornes supérieures.* L'objectif est d'appliquer le Théorème 3.2.2. Prenons  $\nu = 1$ ,

$$\hat{\alpha}_{j,k} = \int_0^1 \phi_{j,k}(t)v(t)dY(t) \quad \text{et} \quad \hat{\beta}_{j,k} = \int_0^1 \psi_{j,k}(t)v(t)dY(t).$$

On a

$$\hat{\beta}_{j,k} - \beta_{j,k} = n^{-1/2} \int_0^1 \psi_{j,k}(t)v(t)dW(t) \sim \mathcal{N}(0, n^{-1} \int_0^1 |\psi_{j,k}(t)|^2 v(t)^2 dt).$$

Par l'hypothèse de bornitude faite sur  $v$ , il vient

$$\int_0^1 |\psi_{j,k}(t)|^2 v(t)^2 dt \leq \|v\|_\infty^2 \int_0^1 |\psi_{j,k}(t)|^2 dt = \|v\|_\infty^2.$$

Or si  $M \sim \mathcal{N}(0, \sigma^2)$  avec  $0 < \sigma \leq \gamma$  alors,

- pour tout  $v \geq 1$ , il existe une constante  $C > 0$  telle que  $\mathbb{E}(|M|^v) \leq C\gamma^v$ ,
- pour tout  $\lambda > 0$ , on a  $\mathbb{P}(|M| \geq \lambda) \leq 2 \exp(-\lambda^2/(2\gamma^2))$ .

L'hypothèse (B1) découle du premier point. En prenant  $\mu$  suffisamment grand, l'hypothèse (B2) découle du second. Le Théorème 3.2.2 nous permet de conclure.

- *Bornes inférieures.* L'objectif est d'appliquer le Théorème 3.2.3. Soient  $h_i, h_l : [0, 1] \rightarrow \mathbb{R}$  des fonctions bornées supérieurement telles que  $h_i \neq h_l$  et  $\mathbb{P}_{h_i}^n \ll \mathbb{P}_{h_l}^n$ .

### 3.5. DÉMONSTRATIONS

---

Le Théorème de Girsanov nous assure que le rapport de vraisemblance  $\Lambda(\mathbb{P}_{h_i}^n, \mathbb{P}_{h_l}^n)$  s'écrit comme

$$\Lambda(\mathbb{P}_{h_i}^n, \mathbb{P}_{h_l}^n) = \exp\left(n \int_0^1 (h_i(t) - h_l(t))v(t)^{-1} dY(t) - 2^{-1}n \int_0^1 (h_i^2(t) - h_l^2(t))v(t)^{-2} dt\right).$$

Sous la loi  $\mathbb{P}_{h_i}^n$ , on a donc

$$\Lambda(\mathbb{P}_{h_i}^n, \mathbb{P}_{h_l}^n) = \exp\left(2^{-1}n \int_0^1 |h_i(t) - h_l(t)|^2 v(t)^{-2} dt + \sqrt{n} \int_0^1 (h_i(t) - h_l(t))v(t)^{-1} dW(t)\right).$$

Par conséquent

$$\mathbb{E}_{h_i}^n(\log(\Lambda(\mathbb{P}_{h_i}^n, \mathbb{P}_{h_l}^n))) = 2^{-1}n \int_0^1 |h_i(t) - h_l(t)|^2 v(t)^{-2} dt. \quad (3.8)$$

Étudions les bornes inférieures suivant le signe de  $\epsilon = \pi s + 2^{-1}(\pi - p)$ .

– Si  $\epsilon > 0$ .

– *Hypothèse (C1)*. Soit  $j_3$  un entier tel que

$$c_0 n^{1/(2s+1)} \leq 2^{j_3} \leq C_0 n^{1/(2s+1)},$$

où  $c_0$  et  $C_0$  désignent deux constantes choisies a posteriori. Pour toute suite  $\varepsilon = (\varepsilon_k)_{k \in \{0, \dots, 2^{j_3}-1\}} \in \{0, 1\}^{2^{j_3}}$ , on définit  $h_\varepsilon : [0, 1] \rightarrow \mathbb{R}$  par

$$h_\varepsilon(x) = L_* 2^{-j_3(s+1/2)} \sum_{k=0}^{2^{j_3}-1} \varepsilon_k \psi_{j_3, k}(x),$$

où  $L_* > 0$  est une constante choisie a posteriori. Les coefficients d'ondelettes de  $h_\varepsilon$  valent

$$\beta_{j, k} = \int_0^1 h_\varepsilon(x) \psi_{j, k}(x) dx = \begin{cases} L_* 2^{-j_3(s+1/2)} \varepsilon_k, & \text{si } j = j_3, \\ 0, & \text{sinon,} \end{cases}$$

d'où  $2^{j_3(s+1/2)} (2^{-j_3} \sum_{k=0}^{2^{j_3}-1} |\beta_{j_3, k}|^\pi)^{1/\pi} = L_*$ . Avec un choix convenable de  $L_*$ , on a  $h_\varepsilon \in B_{\pi, r}^s(L)$ .

Maintenant, on utilise le Théorème de Varshamov-Gilbert décrit par Tsybakov (2004, Lemme 2.7). Celui-ci justifie l'existence d'un sous-ensemble  $E_{j_3} = \{\varepsilon^{(0)}, \dots, \varepsilon^{(T_{j_3})}\}$  de  $\{0, 1\}^{2^{j_3}}$  et de deux constantes  $c, \alpha \in ]0, 1[$  telles que, pour tout  $u \neq v$ ,  $u, v \in \{0, \dots, T_{j_3}\}$ , on a

$$\sum_{k=0}^{2^{j_3}-1} |\varepsilon_k^{(u)} - \varepsilon_k^{(v)}| \geq c 2^{j_3} \quad \text{et} \quad T_{j_3} = \exp(\alpha 2^{j_3}).$$

Grâce à une propriété élémentaire de la base d'ondelettes en norme  $\mathbb{L}^p$  (2.6), pour  $u \neq v$  et  $u, v \in \{0, \dots, T_{j_3}\}$ , on obtient

$$\|h_{\varepsilon(u)} - h_{\varepsilon(v)}\|_p \geq c 2^{j_3/2} 2^{-j_3(s+1/2)} (2^{-j_3} \sum_{k=0}^{2^{j_3}-1} |\varepsilon_k^{(u)} - \varepsilon_k^{(v)}|)^{1/p} \geq c 2^{-j_3 s} = 2v.$$

La famille de fonctions  $h_{\varepsilon(\cdot)}$  satisfait donc l'hypothèse (C1) avec  $m = \exp(\alpha 2^{j_3})$  et  $v \asymp 2^{-j_3 s}$ .

– *Hypothèse (C2)*. Soit  $u \neq v$  avec  $u, v \in \{0, \dots, T_{j_3}\}$ . Pour tout  $t \in [0, 1]$ , la propriété de concentration (2.5) entraîne

$$|h_{\varepsilon(u)}(t) - h_{\varepsilon(v)}(t)| \leq C 2^{-j_3(s+1/2)} \sum_{k=0}^{2^{j_3}-1} |\psi_{j_3,k}(t)| \leq C 2^{-j_3(s+1/2)} 2^{j_3/2} = C 2^{-j_3 s}.$$

Par conséquent, en utilisant la relation (3.8), il vient

$$\begin{aligned} \mathbb{E}_{h_{\varepsilon(u)}}^n (\log(\Lambda(\mathbb{P}_{h_{\varepsilon(u)}}^n, \mathbb{P}_{h_{\varepsilon(v)}}^n))) &= 2^{-1} n \int_0^1 |h_{\varepsilon(u)}(t) - h_{\varepsilon(v)}(t)|^2 v(t)^{-2} dt \\ &\leq C n 2^{-2j_3 s} \int_0^1 v(t)^{-2} dt \leq C c_0^{-2s+1} 2^{j_3}. \end{aligned}$$

L'hypothèse (C2) est donc vérifiée en prenant  $c_0$  suffisamment grand.

En appliquant le Théorème 3.2.3 avec

$$v \asymp 2^{-j_3 s} \asymp n^{-\alpha_1}, \quad \alpha_1 = s/(2s+1),$$

on prouve l'existence d'une constante  $c > 0$  telle que

$$\mathcal{R}_{n,p}(B_{\pi,r}^s(L)) \geq c n^{-\alpha_1 p}.$$

On a la borne inférieure désirée pour  $\epsilon > 0$ .

– Si  $\epsilon \leq 0$ .

– *Hypothèse (C1)*. Soit  $j_4$  un entier vérifiant

$$c_0 (n/\log n)^{1/(2(s-1/\pi)+1)} \leq 2^{j_4} \leq C_0 (n/\log n)^{1/(2(s-1/\pi)+1)},$$

où  $c_0$  et  $C_0$  désignent deux constantes choisies a posteriori. On définit  $h_\varrho : [0, 1] \rightarrow \mathbb{R}$  par

$$\begin{cases} h_\varrho(x) = L_* 2^{-j_4(s+1/2-1/\pi)} \psi_{j_4,\varrho}(x), & \varrho \in \{0, \dots, 2^{j_4} - 1\}, \\ h_{2^{j_4}}(x) = 0, \end{cases} \quad (3.9)$$

### 3.5. DÉMONSTRATIONS

---

où  $L_* > 0$  est une constante choisie a posteriori. Comme les coefficients d'ondelettes de  $h_\varrho$  valent

$$\beta_{j,k} = \int_0^1 h_\varrho(x) \psi_{j,k}(x) dx = \begin{cases} L_* 2^{-j_4(s+1/2-1/\pi)}, & \text{si } j = j_4 \quad \text{et} \quad k = \varrho, \\ 0, & \text{sinon,} \end{cases}$$

on a  $2^{j_4(s+1/2)}(2^{-j_4}|\beta_{j_4,\varrho}|^\pi)^{1/\pi} = L_*$ . Avec un choix convenable de  $L_*$ , on a  $h_\varrho \in B_{\pi,r}^s(L)$  pour tout  $\varrho \in \{0, \dots, 2^{j_4}\}$ .

De plus, en utilisant une propriété élémentaire de la base d'ondelettes en norme  $\mathbb{L}^p$  (2.6), pour tout  $\varrho \neq \varrho'$  avec  $\varrho, \varrho' \in \{0, \dots, 2^{j_4}\}$ , on a

$$\|h_\varrho - h_{\varrho'}\|_p \geq c 2^{-j_4(s+1/2-1/\pi)} 2^{j_4(1/2-1/p)} = c 2^{-j_4(s-1/\pi+1/p)} = 2v.$$

La famille de fonctions  $h_\cdot$  satisfait donc l'hypothèse (C1) avec  $m = 2^{j_4}$  et  $v \asymp 2^{-j_4(s-1/\pi+1/p)}$ .

– *Hypothèse (C2)*. La relation (3.8) et la propriété de concentration (2.5) entraînent

$$\begin{aligned} \sum_{k=0}^{2^{j_4}-1} \mathbb{E}_{h_\varrho}^n(\log(\Lambda(\mathbb{P}_{h_\varrho}^n, \mathbb{P}_{h_{2^{j_4}}}^n))) &= 2^{-1}n \sum_{\varrho=0}^{2^{j_4}-1} \int_0^1 |h_\varrho(t) - h_{2^{j_4}}(t)|^2 v(t)^{-2} dt \\ &\leq Cn 2^{-2j_4(s+1/2-1/\pi)} \int_0^1 v(t)^{-2} \sum_{\varrho=0}^{2^{j_4}-1} |\psi_{j_4,\varrho}(t)|^2 dt \\ &\leq Cn 2^{-2j_4(s+1/2-1/\pi)} 2^{j_4} \|1/v\|_2^2 \\ &\leq Cc_0^{-2(s+1/2-1/\pi)} \log n 2^{j_4}. \end{aligned}$$

Or, pour  $n$  large, on a

$$\log(2^{j_4}) \geq (\log n - \log(\log n) + C)/(2(s+1/2-1/\pi)) \geq c \log n.$$

Par un choix convenable de  $c_0$ , on déduit l'existence d'une constante  $p_0 \in ]0, 8^{-1}[$  vérifiant

$$2^{-j_4} \sum_{\varrho=0}^{2^{j_4}-1} \mathbb{E}_{h_\varrho}^n(\log(\Lambda(\mathbb{P}_{h_\varrho}^n, \mathbb{P}_{h_{2^{j_4}}}^n))) \leq p_0 \log(2^{j_4}).$$

L'hypothèse (C2) est vérifiée. Le Théorème 3.2.3 appliqué avec

$$v \asymp 2^{-j_4(s-1/\pi+1/p)} \asymp (\log n/n)^{\alpha_2}, \quad \alpha_2 = (s-1/\pi+1/p)/(2(s-1/\pi)+1),$$

justifie l'existence d'une constante  $c > 0$  telle que

$$\mathcal{R}_{n,p}(B_{\pi,r}^s(L)) \geq c (\log n/n)^{\alpha_2 p}.$$



On a la borne inférieure désirée pour  $\epsilon \leq 0$ . La preuve du Théorème 3.3.1 est complète.  $\square$

**PREUVE DU THÉORÈME 3.3.2.** Puisque la borne inférieure du Théorème 3.3.1 a été obtenue sous l'unique hypothèse  $1/v \in \mathbb{L}^2([0, 1])$ , elle reste vraie dans le cadre statistique du Théorème 3.3.2. Par conséquent, seule la borne supérieure est à montrer.

• *Borne supérieure.* Le but est d'appliquer le Théorème 3.2.1. Seule l'hypothèse (A) est à satisfaire. Pour tout entier  $j_1 \geq \tau$ , considérons l'estimateur

$$\hat{\alpha}_{j_1,k} = \int_0^1 \phi_{j_1,k}(t) v(t) dY(t).$$

On a

$$\hat{\alpha}_{j_1,k} - \alpha_{j_1,k} = n^{-1/2} \int_0^1 \phi_{j_1,k}(t) v(t) dW(t) \sim \mathcal{N}(0, n^{-1} \rho_{j_1,k}^2),$$

où  $\rho_{j_1,k}$  est défini par

$$\rho_{j_1,k} = \left( \int_0^1 |\phi_{j_1,k}(t)|^2 v(t)^2 dt \right)^{1/2}. \quad (3.10)$$

Par une inégalité gaussienne élémentaire, on a

$$\sum_{k=0}^{2^{j_1}-1} \mathbb{E}_f^n(|\hat{\alpha}_{j_1,k} - \alpha_{j_1,k}|^p) \leq C n^{-p/2} \sum_{k=0}^{2^{j_1}-1} \rho_{j_1,k}^p. \quad (3.11)$$

Étudions séparément le cas où  $1 \leq p \leq 2$  et le cas où  $p > 2$ .

– Si  $1 \leq p \leq 2$ . L'inégalité de Hölder, la propriété de concentration (2.5) et l'hypothèse (3.5) impliquent

$$\begin{aligned} \sum_{k=0}^{2^{j_1}-1} \rho_{j_1,k}^p &\leq 2^{j_1(1-p/2)} \left( \sum_{k=0}^{2^{j_1}-1} \int_0^1 v(t)^2 |\phi_{j_1,k}(t)|^2 dt \right)^{p/2} \\ &\leq C 2^{j_1(1-p/2)} (2^{j_1} \int_0^1 v(t)^2 dt)^{p/2} \leq C 2^{j_1}. \end{aligned}$$

– Si  $p > 2$ . Par l'inégalité de Hölder appliquée à la mesure  $d\omega = |\phi_{j_1,k}(t)|^2 dt$ , la propriété de concentration (2.5) et la condition (3.5), il vient

$$\sum_{k=0}^{2^{j_1}-1} \rho_{j_1,k}^p \leq \int_0^1 v^p(t) \sum_{k=0}^{2^{j_1}-1} |\phi_{j_1,k}(t)|^2 dt \leq C \|v\|_p^p 2^{j_1} \leq C 2^{j_1}.$$

Par conséquent, pour tout  $p \geq 1$ , on a

$$\sum_{k=0}^{2^{j_1}-1} \mathbb{E}_f^n(|\hat{\alpha}_{j_1,k} - \alpha_{j_1,k}|^p) \leq C n^{-p/2} 2^{j_1}.$$

### 3.5. DÉMONSTRATIONS

---

Le Théorème 3.2.1 s'applique avec  $\eta = 1$  et  $j_1$  un entier défini par

$$2^{j_1} \asymp n^{1/(2s+1)}.$$

Ce qui donne la vitesse de convergence

$$\varphi_n = n^{-p/2} 2^{j_1 p/2} + 2^{-j_1 s p} \asymp n^{-\alpha_1 p}, \quad \alpha_1 = s/(2s+1).$$

La preuve du Théorème 3.3.2 est terminée.  $\square$

PREUVE DE LA PROPOSITION 3.3.1. Considérons la fonction  $v$  définie par

$$v(t) = t^{-\sigma/2}, \quad \sigma > 2p^{-1}.$$

L'idée principale est d'utiliser le Théorème 3.2.3 pour exhiber une borne inférieure strictement supérieure à  $n^{-\alpha_1 p}$  et  $(\log n/n)^{\alpha_2 p}$ .

– *Hypothèse (C1)*. Soit  $j_3$  un entier tel que

$$c_0 n^{1/(2(s-1/\pi)+\sigma+1)} \leq 2^{j_3} \leq C_0 n^{1/(2(s-1/\pi)+\sigma+1)},$$

où  $c_0$  et  $C_0$  désignent deux constantes choisies a posteriori. On définit  $h_{-1}, h_1 : [0, 1] \rightarrow \mathbb{R}$  par

$$\begin{cases} h_1(x) = L_* 2^{-j_3(s-1/\pi+1/2)} \psi_{j_3, N}(x), \\ h_{-1}(x) = -L_* 2^{-j_3(s-1/\pi+1/2)} \psi_{j_3, N}(x), \end{cases} \quad (3.12)$$

où  $L_* > 0$  est une constante choisie a posteriori. Comme les coefficients d'ondelettes de  $h_1$  et  $h_{-1}$  valent

$$\beta_{j,k} = \int_0^1 h_{\pm 1}(x) \psi_{j,k}(x) dx = \begin{cases} \pm L_* 2^{-j_3(s+1/2-1/\pi)}, & \text{si } j = j_3 \text{ et } k = N, \\ 0, & \text{sinon,} \end{cases}$$

on a  $2^{j_3(s+1/2)}(2^{-j_3} |\beta_{j_3, N}|^\pi)^{1/\pi} = L_*$ . Avec un choix convenable de  $L_*$ , les fonctions  $h_1$  et  $h_{-1}$  appartiennent donc à  $B_{\pi, r}^s(L)$ .

En utilisant une propriété élémentaire de la base d'ondelettes en norme  $\mathbb{L}^p$  (2.6), on a

$$\|h_1 - h_{-1}\|_p \geq c 2^{j_3(1/2-1/p)} 2^{-j_3(s+1/2-1/\pi)} = c 2^{-j_3(s-1/\pi+1/p)} = 2v.$$

Les fonctions  $h_1$  et  $h_{-1}$  vérifient donc l'hypothèse (C1) avec  $m = 2$  et  $v \asymp 2^{-j_3(s-1/\pi+1/p)}$ .

– *Hypothèse (C2)*. Comme  $\text{Supp}(\psi_{j_3,N}) \subseteq [2^{-j_3}, (2N)2^{-j_3}]$  et  $t^\sigma$  est croissante pour  $\sigma > 0$ , la relation (3.8) entraîne

$$\begin{aligned} \mathbb{E}_{h_1}^n(\log(\Lambda(\mathbb{P}_{h_1}^n, \mathbb{P}_{h_{-1}}^n))) &= 2^{-1}n \int_0^1 |h_1(t) - h_{-1}(t)|^2 v(t)^{-2} dt \\ &= C 2^{-2j_3(s+1/2-1/\pi)} n \int_{2^{-j_3}}^{(2N)2^{-j_3}} t^\sigma |\psi_{j_3,N}(t)|^2 dt \\ &\leq C 2^{-2j_3(s+1/2-1/\pi)} n 2^{-j_3\sigma} \int_{2^{-j_3}}^{(2N)2^{-j_3}} |\psi_{j_3,N}(t)|^2 dt \\ &= C 2^{-2j_3(s+1/2-1/\pi)} n 2^{-j_3\sigma} \leq C c_0^{-2(s-1/\pi)+\sigma+1}. \end{aligned}$$

En prenant  $c_0$  suffisamment grand, on montre l'existence d'une constante  $p_0 \in ]0, 8^{-1}[$  telle que

$$\mathbb{E}_{h_1}^n(\log(\Lambda(\mathbb{P}_{h_1}^n, \mathbb{P}_{h_{-1}}^n))) \leq p_0 \log(2).$$

L'hypothèse (C2) est vérifiée.

Le Théorème 3.2.3 s'applique avec

$$v \asymp 2^{-j_3(s-1/\pi+1/p)} \asymp n^{-\alpha_3}, \quad \alpha_3 = (s - 1/\pi + 1/p)/(2(s - 1/\pi) + 1 + \sigma).$$

Cela justifie l'existence d'une constante  $c > 0$  telle que

$$\mathcal{R}_{n,p}(B_{\pi,r}^s(L)) \geq c n^{-\alpha_3 p}.$$

Comme  $\alpha_3 < \alpha_1 \wedge \alpha_2$  pour  $\sigma$  suffisamment grand, la preuve de la Proposition 3.3.1 est terminée.  $\square$

**PREUVE DU THÉORÈME 3.3.3.** La borne inférieure a été calculée dans la preuve de la Proposition 3.3.1 ci-dessus. Seule la borne supérieure demande une étude.

• *Borne supérieure.* L'objectif est d'appliquer le Théorème 3.1. Il suffit donc de satisfaire l'hypothèse (A). Pour tout entier  $j_1 \geq \tau$ , considérons l'estimateur

$$\hat{\alpha}_{j_1,k} = \int_0^1 \phi_{j_1,k}(t) t^{-\sigma/2} dY(t).$$

En procédant comme pour la borne supérieure du Théorème 3.3.2, il vient

$$\sum_{k=0}^{2^{j_1}-1} \mathbb{E}_f^n(|\hat{\alpha}_{j_1,k} - \alpha_{j_1,k}|^p) \leq C n^{-p/2} \sum_{k=0}^{2^{j_1}-1} \rho_{j_1,k}^p,$$

où

$$\rho_{j_1,k} = \left( \int_0^1 |\phi_{j_1,k}(t)|^2 t^{-\sigma} dt \right)^{1/2}. \quad (3.13)$$

### 3.5. DÉMONSTRATIONS

---

Les éléments de la base d'ondelettes  $\zeta$  construite par Cohen et al. (1993) vérifient

$$Supp(\phi_{j,k}) = Supp(\psi_{j,k}) \subseteq \begin{cases} [0, (2N-1)2^{-j}], & \text{si } k \in \{0, \dots, N-1\}, \\ [N, 2^j - N - 1], & \text{si } k \in \{N, \dots, 2^j - N - 1\}, \\ [1 - (2N-1)2^{-j}, 1], & \text{si } k \in \{2^j - N, \dots, 2^j - 1\}. \end{cases}$$

C'est pourquoi on considère la décomposition suivante :

$$\sum_{k=0}^{2^{j_1}-1} \rho_{j_1,k}^p = \sum_{k=0}^{N-1} \rho_{j_1,k}^p + \sum_{k=N}^{2^{j_1}-N-1} \rho_{j_1,k}^p + \sum_{k=2^{j_1}-N}^{2^{j_1}-1} \rho_{j_1,k}^p.$$

Pour le premier terme, on a

$$\sum_{k=0}^{N-1} \rho_{j_1,k}^p \leq N \|\phi\|_\infty^p 2^{j_1 p/2} \left( \int_0^{(2N-1)2^{-j_1}} t^{-\sigma} dt \right)^{p/2} = C 2^{j_1 p/2} (2^{j_1(\sigma-1)p/2}) = C 2^{j_1 \sigma p/2}.$$

Comme  $t^{-\sigma}$  est décroissante pour  $\sigma > 0$ ,  $\sigma p/2 > 1$  par hypothèse et  $\sum_{k=1}^\infty k^{-\beta} < \infty$  pour  $\beta > 1$ , on a

$$\begin{aligned} \sum_{k=N}^{2^{j_1}-N-1} \rho_{j_1,k}^p &= \sum_{k=N}^{2^{j_1}-N-1} \left( \int_{(k-N+1)2^{-j_1}}^{(k+N)2^{-j_1}} |\phi_{j_1,k}(t)|^2 t^{-\sigma} dt \right)^{p/2} \\ &\leq 2^{j_1 \sigma p/2} \sum_{k=N}^{2^{j_1}-N-1} (k - N + 1)^{-\sigma p/2} \leq C 2^{j_1 \sigma p/2}. \end{aligned}$$

En utilisant l'inégalité élémentaire  $1 - (1-x)^\alpha \leq x^\alpha$ , pour  $0 < \alpha < 1$  et  $x \in [0, 1/2]$ , on trouve

$$\begin{aligned} \sum_{k=2^{j_1}-N}^{2^{j_1}-1} \rho_{j_1,k}^p &\leq N \|\phi\|_\infty^p 2^{j_1 p/2} \left( \int_{1-(2N-1)2^{-j_1}}^1 t^{-\sigma} dt \right)^{p/2} \\ &\leq C 2^{j_1 p/2} (1 - (1 - (2N-1)2^{-j_1})^{-\sigma+1})^{p/2} \leq C 2^{j_1 \sigma p/2}. \end{aligned}$$

En combinant les inégalités précédentes, on aboutit à

$$\sum_{k=0}^{2^{j_1}-1} \mathbb{E}_f^n(|\hat{\alpha}_{j_1,k} - \alpha_{j_1,k}|^p) \leq C n^{-p/2} 2^{j_1 \sigma p/2}.$$

Le Théorème 3.2.1 s'applique avec  $\eta = \sigma p/2$  et  $j_1$  un entier défini par

$$2^{j_1} \asymp n^{1/(2(s-1/\pi)+\sigma+1)}.$$

Ce qui donne la vitesse de convergence

$$\begin{aligned} \varphi_n &= n^{-p/2} 2^{j_1(p/2-1+\sigma p/2)} + 2^{-j_1 \sigma p} 2^{j_1(p-\pi)/\pi} \asymp n^{-\alpha_3 p}, \\ \alpha_3 &= (s - 1/\pi + 1/p)/(2(s - 1/\pi) + \sigma + 1). \end{aligned}$$

La preuve du Théorème (3.3.3) est terminée. □

### 3.5.2 Preuves des résultats de la Section 3.4

Trois inégalités probabilistes utilisées dans cette section sont rappelées ci-dessous.

- *Inégalité de Rosenthal.* Soit  $p \in [2, \infty[$ . Soit  $(W_1, \dots, W_n)$  des variables indépendantes vérifiant  $\mathbb{E}(W_i) = 0$  et  $\mathbb{E}(|W_i|^p) < \infty$ ,  $i = 1, \dots, n$ . Alors il existe une constante  $C > 0$  telle que

$$\mathbb{E}(|\sum_{i=1}^n W_i|^p) \leq C(\sum_{i=1}^n \mathbb{E}(|W_i|^p) + (\sum_{i=1}^n \mathbb{E}(|W_i|^2))^{p/2}).$$

- *Inégalité de Hoeffding.* Soit  $(a_i)_{i \in \mathbb{N}}$  et  $(b_i)_{i \in \mathbb{N}}$  deux suites réelles positives. Soit  $(W_1, \dots, W_n)$  des variables indépendantes vérifiant  $a_i \leq W_i \leq b_i$  et  $\mathbb{E}(W_i) = 0$ ,  $i = 1, \dots, n$ . Alors, pour tout  $\lambda > 0$ , on a

$$\mathbb{P}(\sum_{i=1}^n W_i \geq \lambda) \leq \exp(-2\lambda^2 / (\sum_{i=1}^n (b_i - a_i)^2)).$$

- *Inégalité de Bernstein.* Soit  $(W_1, \dots, W_n)$  des variables indépendantes vérifiant  $W_i \leq C$ ,  $\mathbb{E}(W_i) \leq 0$ ,  $i = 1, \dots, n$  et  $b_n^2 = \sum_{i=1}^n \mathbb{E}(|W_i|^2) < \infty$ . Alors, pour tout  $\lambda > 0$ , on a

$$\mathbb{P}(\sum_{i=1}^n W_i \geq \lambda) \leq \exp(-\lambda^2 / (2(b_n^2 + \lambda C/3))).$$

PREUVE DU THÉORÈME 3.4.1. Étudions séparément les bornes inférieures et les bornes supérieures.

- *Bornes supérieures.* L'objectif est d'appliquer le Théorème (3.2.2). Posons  $\nu = 2^{-1}$ ,

$$\hat{\alpha}_{j,k} = n^{-1} \sum_{i=1}^n Y_i g(X_i)^{-1} \phi_{j,k}(X_i) \quad \text{et} \quad \hat{\beta}_{j,k} = n^{-1} \sum_{i=1}^n Y_i g(X_i)^{-1} \psi_{j,k}(X_i). \quad (3.14)$$

Montrons que les hypothèses (B1) et (B2) sont vérifiées.

- *Hypothèse (B1).* Tout d'abord, remarquons que

$$\mathbb{E}_f^n(\hat{\beta}_{j,k}) = \mathbb{E}_f^n(f(X_1)\psi_{j,k}(X_1)g(X_1)^{-1}) = \int_0^1 f(t)\psi_{j,k}(t)g(t)g(t)^{-1}dt = \beta_{j,k}.$$

L'inégalité de Rosenthal appliquée aux variables aléatoires i.i.d  $(W_1, \dots, W_n)$  avec  $W_i = Y_i\psi_{j,k}(X_i)g(X_i)^{-1} - \beta_{j,k}$  justifie l'existence d'une constante  $C > 0$  telle que

$$\mathbb{E}_f^n(|\hat{\beta}_{j,k} - \beta_{j,k}|^{2p}) \leq C(n^{1-2p}Z_{2p} + n^{-p}(Z_2)^p), \quad (3.15)$$

où

$$Z_a = \mathbb{E}_f^n(|W_1|^a), \quad a \geq 2.$$

### 3.5. DÉMONSTRATIONS

---

Par une inégalité élémentaire de convexité et l'inégalité de Hölder, on obtient

$$\begin{aligned} Z_a &\leq 2^{a-1} (\mathbb{E}_f^n(|Y_i \psi_{j,k}(X_i) g(X_i)^{-1}|^a) + \mathbb{E}_f^n(|Y_i \psi_{j,k}(X_i) g(X_i)^{-1}|^a)) \\ &\leq C \mathbb{E}_f^n(|Y_i \psi_{j,k}(X_i) g(X_i)^{-1}|^a) \leq C(Z_{a,1} + Z_{a,2}), \end{aligned} \quad (3.16)$$

où

$$Z_{a,1} = \mathbb{E}_f^n(|f(X_1) \psi_{j,k}(X_1) g(X_1)^{-1}|^a) \quad \text{et} \quad Z_{a,2} = \mathbb{E}_f^n(|z_1 \psi_{j,k}(X_1) g(X_1)^{-1}|^a).$$

Par les hypothèses de bornitude faites sur  $f$  et  $g$  et le fait que  $2^j \leq 2^{j_2} \leq n$ , il vient

$$\begin{aligned} Z_{a,1} \vee Z_{a,2} &\leq \|1/g\|_\infty^a (\|f\|_\infty^a \vee \mathbb{E}_f^n(|z_1|^a)) \int_0^1 |\psi_{j,k}(x)|^a dx \\ &\leq C \|\psi\|_\infty^{a-2} 2^{j(a/2-1)} \int_0^1 |\psi_{j,k}(t)|^2 dt \leq C n^{a/2-1}. \end{aligned} \quad (3.17)$$

En combinant les inégalités (3.15), (3.16) et (3.17), on obtient

$$\mathbb{E}_f^n(|\hat{\beta}_{j,k} - \beta_{j,k}|^{2p}) \leq C(n^{1-2p} n^{p-1} + n^{-p}) \leq C n^{-p}.$$

– *Hypothèse (B2)*. On utilise quelques techniques de majorations similaires à celles développées par Kerkycharian et Picard (2005, Subsection 9.1). On rappelle que  $a, j \in \{j_1, \dots, j_2\}$ . Par une décomposition élémentaire, il vient

$$\mathbb{P}_f^n(|\hat{\beta}_{j,k} - \beta_{j,k}| \geq 2^{-1} \mu \sqrt{(a/n)}) \leq S_1 + S_2,$$

où

$$S_1 = \mathbb{P}_f^n(|n^{-1} \sum_{i=1}^n (\psi_{j,k}(X_i) f(X_i) g(X_i)^{-1} - \beta_{j,k})| \geq 4^{-1} \mu \sqrt{(a/n)})$$

et

$$S_2 = \mathbb{P}_f^n(|n^{-1} \sum_{i=1}^n \psi_{j,k}(X_i) g(X_i)^{-1} z_i| \geq 4^{-1} \mu \sqrt{(a/n)}).$$

Étudions les majorations des termes  $S_1$  et  $S_2$ .

– *Majoration de  $S_1$* . Par les hypothèses de bornitude faites sur  $f$  et  $g$  et  $\|\psi_{j,k}\|_\infty \leq \|\psi\|_\infty 2^{j/2}$ , on a

$$|\psi_{j,k}(X_1) f(X_1) g(X_1)^{-1} - \beta_{j,k}| \leq |\psi_{j,k}(X_1) f(X_1) g(X_1)^{-1}| + |\beta_{j,k}| \leq C 2^{j/2}$$

et

$$\mathbb{E}_f^n(|\psi_{j,k}(X_1) f(X_1) g(X_1)^{-1} - \beta_{j,k}|^2) \leq C \mathbb{E}_f^n(|\psi_{j,k}(X_1) f(X_1) g(X_1)^{-1}|^2) \leq C.$$

Pour tout  $j \in \{j_s+1, \dots, j_2\}$ , l'inégalité de Bernstein appliquée aux variables aléatoires i.i.d.  $(\psi_{j,k}(X_1) f(X_1) g(X_1)^{-1}, \dots, \psi_{j,k}(X_n) f(X_n) g(X_n)^{-1})$  nous donne

$$\begin{aligned} S_1 &\leq 2 \exp(-Cn(\mu \sqrt{(a/n)})^2 / (1 + 2^{j/2} 4^{-1} \mu \sqrt{(a/n)})) \\ &\leq 2 \exp(-Cn(\mu \sqrt{(a/n)})^2 / (1 + 2^{j_2/2} 4^{-1} \mu \sqrt{(\log n/n)})) \\ &\leq 2 \exp(-C\mu^2 a). \end{aligned}$$

Par conséquent, il existe deux constantes  $\mu > 0$  et  $C > 0$  telles que

$$S_1 \leq C 2^{-2ap}. \quad (3.18)$$

– *Majoration de  $S_2$ .* Conditionnellement à  $(X_1, \dots, X_n)$ ,  $n^{-1} \sum_{i=1}^n \psi_{j,k}(X_i) g(X_i)^{-1} z_i \sim \mathcal{N}(0, n^{-2} \sum_{i=1}^n |\psi_{j,k}(X_i)|^2 g(X_i)^{-2})$ . Par l'hypothèse de minoration faite sur  $g$ , remarquons que

$$n^{-2} \sum_{i=1}^n |\psi_{j,k}(X_i)|^2 g(X_i)^{-2} \leq \|1/g\|_\infty n^{-2} \sum_{i=1}^n |\psi_{j,k}(X_i)|^2 g(X_i)^{-1}.$$

Considérons l'ensemble  $\mathcal{F}_\mu$  défini par

$$\mathcal{F}_\mu = \{ |n^{-1} \sum_{i=1}^n |\psi_{j,k}(X_i)|^2 g(X_i)^{-1} - 1| \geq \mu \}$$

et la quantité  $\mathcal{Q}$  définie par

$$\mathcal{Q} = 2 \exp(-\mu^2 a / (32 \|1/g\|_\infty n^{-1} \sum_{i=1}^n |\psi_{j,k}(X_i)|^2 g(X_i)^{-1})).$$

Par une inégalité gaussienne élémentaire et la majoration  $\mathcal{Q} \leq 2$ , il vient

$$\begin{aligned} S_2 &= \mathbb{E}_f^n(\mathbb{P}_f^n(|n^{-1} \sum_{i=1}^n \psi_{j,k}(X_i) g(X_i)^{-1} z_i| \geq 4^{-1} \mu \sqrt{a/n} |X_1, \dots, X_n)) \\ &\leq \mathbb{E}_f^n(\mathcal{Q} 1_{\mathcal{F}_\mu}) + \mathbb{E}_f^n(\mathcal{Q} 1_{\mathcal{F}_\mu^c}) \leq 2 \mathbb{P}_f^n(\mathcal{F}_\mu) + \mathbb{E}_f^n(\mathcal{Q} 1_{\mathcal{F}_\mu^c}). \end{aligned}$$

Puisque sur l'ensemble  $\mathcal{F}_\mu^c$ , on a  $\mathcal{Q} \leq 2 \exp(-\mu^2 a / (32 \|1/g\|_\infty (1 + \mu)))$ , l'inégalité élémentaire  $\mathbb{P}_f^n(\mathcal{F}_\mu^c) \leq 1$  donne

$$S_2 \leq 2(\mathbb{P}_f^n(\mathcal{F}_\mu) + \exp(-\mu^2 a / (32 \|1/g\|_\infty (1 + \mu)))). \quad (3.19)$$

Puisque les variables aléatoires  $(|\psi_{j,k}(X_1)|^2 g(X_1)^{-1}, \dots, |\psi_{j,k}(X_n)|^2 g(X_n)^{-1})$  sont i.i.d avec

$$|\psi_{j,k}(X_i)|^2 g(X_i)^{-1} \leq \|1/g\|_\infty \|\psi\|_\infty^2 2^j \quad \text{et} \quad \mathbb{E}_f^n(|\psi_{j,k}(X_1)|^2 g(X_1)^{-1}) = 1,$$

pour tout  $j \in \{j_s + 1, \dots, j_2\}$  et  $\mu$  suffisamment grand, l'inégalité de Hoeffding entraîne

$$\begin{aligned} \mathbb{P}_f^n(\mathcal{F}_\mu) &\leq 2 \exp(-2n\mu^2 2^{-2j} / (\|\psi\|_\infty^4 \|1/g\|_\infty^2)) \leq 2 \exp(-2n\mu^2 2^{-2j_2} / (\|\psi\|_\infty^4 \|1/g\|_\infty^2)) \\ &\leq 2n^{-C\mu^2} \leq 22^{-aC\mu^2}. \end{aligned} \quad (3.20)$$

Il découle des inégalités (3.19) et (3.20) l'existence de deux constantes  $\mu > 0$  et  $C > 0$  telles que

$$S_2 \leq C 2^{-2ap}. \quad (3.21)$$


---

### 3.5. DÉMONSTRATIONS

---

En mettant les inégalités (3.18), (3.19), (3.20) et (3.21) bout à bout, on montre l'hypothèse (B2). Le Théorème 3.2.2 permet de conclure.

• *Bornes inférieures.* L'objectif est d'utiliser le Théorème 3.2.3. Pour toutes fonctions  $h_v, h_l : [0, 1] \rightarrow \mathbb{R}$  bornées supérieurement telles que  $h_v \neq h_l$  et  $\mathbb{P}_{h_v}^n \ll \mathbb{P}_{h_l}^n$ , le rapport de vraisemblance associé au modèle s'écrit comme

$$\begin{aligned} \Lambda(\mathbb{P}_{h_v}^n, \mathbb{P}_{h_l}^n) &= \exp(-2^{-1}(\sum_{i=1}^n |Y_i - h_v(X_i)|^2 - \sum_{i=1}^n |Y_i - h_l(X_i)|^2)) \\ &= \exp(-2^{-1}(\sum_{i=1}^n (h_l(X_i) - h_v(X_i))(2Y_i - h_v(X_i) - h_l(X_i)))). \end{aligned}$$

Par un développement élémentaire, sous la probabilité  $\mathbb{P}_{h_v}^n$ , on a

$$\Lambda(\mathbb{P}_{h_v}^n, \mathbb{P}_{h_l}^n) = \exp(2^{-1} \sum_{i=1}^n |h_v(X_i) - h_l(X_i)|^2 + \sum_{i=1}^n (h_v(X_i) - h_l(X_i))z_i).$$

Par indépendance des variables aléatoires et le fait que  $z_1 \sim \mathcal{N}(0, 1)$ , il vient

$$\mathbb{E}_{h_v}^n(\log(\Lambda(\mathbb{P}_{h_v}^n, \mathbb{P}_{h_l}^n))) = 2^{-1}n \int_0^1 |h_v(t) - h_l(t)|^2 g(t) dt. \quad (3.22)$$

Cette identité est similaire à (3.8). La densité  $g$  a juste remplacé la fonction  $1/v^2$ . En procédant comme pour la preuve des bornes inférieures du Théorème 3.3.1, on établit les résultats souhaités. Cela termine la preuve du Théorème 3.4.1.  $\square$

**PREUVE DU THÉORÈME 3.4.2.** L'objectif est d'appliquer le Théorème 3.1. Montrons que l'hypothèse (A) est satisfaite. Considérons l'estimateur

$$\hat{\alpha}_{j_1, k} = n^{-1} \sum_{i=1}^n Y_i g(X_i)^{-1} \psi_{j_1, k}(X_i).$$

Supposons que  $j_1$  est un entier vérifiant  $2^{j_1} \leq Cn$ .

Soit  $a \geq 2$ . L'inégalité de Rosenthal appliquée aux variables aléatoires i.i.d  $(W_1, \dots, W_n)$  avec  $W_i = Y_i g(X_i)^{-1} \phi_{j_1, k}(X_i) - \alpha_{j_1, k}$  justifie l'existence d'une constante  $C > 0$  telle que

$$\mathbb{E}_f^n(|\hat{\alpha}_{j_1, k} - \alpha_{j_1, k}|^a) \leq C(n^{1-a} Z_a + n^{-a/2} (Z_2)^{a/2}), \quad (3.23)$$

où

$$Z_a = \mathbb{E}_f^n(|W_1|^a).$$

En utilisant une inégalité élémentaire de convexité et l'inégalité de Hölder, il vient

$$\begin{aligned} Z_a &\leq 2^{a-1} (\mathbb{E}_f^n(|Y_1 g(X_1)^{-1} \phi_{j_1, k}(X_1)|^a) + \mathbb{E}_f^n(|Y_1 g(X_1)^{-1} \phi_{j_1, k}(X_1)|)^a) \\ &\leq C \mathbb{E}_f^n(|Y_1 g(X_1)^{-1} \phi_{j_1, k}(X_1)|^a) \leq C(Z_{a,1} + Z_{a,2}), \end{aligned} \quad (3.24)$$



où

$$Z_{a,1} = \mathbb{E}_f^n(|f(X_1)g(X_1)^{-1}\phi_{j_1,k}(X_1)|^a) \quad \text{et} \quad Z_{a,2} = \mathbb{E}_f^n(|z_1\phi_{j_1,k}(X_1)g(X_1)^{-1}|^a).$$

Supposons que  $1/g \in \mathbb{L}^{a-1}([0, 1])$ . On a

$$Z_a \leq C(Z_{a,1} \vee Z_{a,2}) \leq C(\|f\|_\infty^a \vee \mathbb{E}_f^n(|z_1|^a)) \int_0^1 |\phi_{j_1,k}(x)|^a g(x)^{1-a} dx.$$

L'inégalité de Hölder appliquée à la mesure  $d\omega = |\phi_{j_1,k}(x)|^2 dx$  implique

$$\begin{aligned} (Z_2)^{a/2} &\leq C((Z_{2,1})^{a/2} \vee (Z_{2,2})^{a/2}) \\ &\leq C(\|f\|_\infty^a \vee \mathbb{E}_f^n(|z_1|^a)) \left( \int_0^1 |\phi_{j_1,k}(x)|^2 g(x)^{-1} dx \right)^{a/2} \\ &\leq C \int_0^1 |\phi_{j_1,k}(x)|^2 g(x)^{-a/2} dx. \end{aligned} \quad (3.25)$$

En mettant les inégalités (3.23), (3.24) et (3.25) ensemble, on obtient

$$\begin{aligned} &\mathbb{E}_f^n(|\hat{\alpha}_{j_1,k} - \alpha_{j_1,k}|^a) \\ &\leq C(n^{1-a} \int_0^1 |\phi_{j_1,k}(x)|^a g(x)^{1-a} dx + n^{-a/2} \int_0^1 |\phi_{j_1,k}(x)|^2 g(x)^{-a/2} dx). \end{aligned} \quad (3.26)$$

Distinguons le cas où  $1 \leq p \leq 2$  et le cas où  $p > 2$ .

– Si  $1 \leq p \leq 2$ . En utilisant l'inégalité de Jensen et l'inégalité (3.26) avec  $a = 2$ , il vient

$$\mathbb{E}_f^n(|\hat{\alpha}_{j_1,k} - \alpha_{j_1,k}|^p) \leq \mathbb{E}_f^n(|\hat{\alpha}_{j_1,k} - \alpha_{j_1,k}|^2)^{p/2} \leq C(n^{-1} \int_0^1 |\phi_{j_1,k}(x)|^2 g(x)^{-1} dx)^{p/2}.$$

Par l'inégalité de Hölder  $l_p$ , la propriété de concentration (2.5) et l'hypothèse (3.7), il vient

$$\begin{aligned} \sum_{k=0}^{2^{j_1}-1} \mathbb{E}_f^n(|\hat{\alpha}_{j_1,k} - \alpha_{j_1,k}|^p) &\leq Cn^{-p/2} 2^{j_1(1-p/2)} \left( \sum_{k=0}^{2^{j_1}-1} \int_0^1 |\phi_{j_1,k}(x)|^2 g(x)^{-1} dx \right)^{p/2} \\ &\leq Cn^{-p/2} 2^{j_1(1-p/2)} \left( 2^{j_1} \int_0^1 g(x)^{-1} dx \right)^{p/2} \leq C2^{j_1} n^{-p/2}. \end{aligned}$$

– Si  $p > 2$ . L'inégalité (3.26) prise avec  $a = p$  donne

$$\begin{aligned} &\sum_{k=0}^{2^{j_1}-1} \mathbb{E}_f^n(|\hat{\alpha}_{j_1,k} - \alpha_{j_1,k}|^p) \\ &\leq C(n^{1-p} \int_0^1 g(x)^{1-p} \sum_{k=0}^{2^{j_1}-1} |\phi_{j_1,k}(x)|^p dx + n^{-p/2} \int_0^1 g(x)^{-p/2} \sum_{k=0}^{2^{j_1}-1} |\phi_{j_1,k}(x)|^2 dx) = U. \end{aligned}$$

### 3.5. DÉMONSTRATIONS

---

La propriété de concentration (2.5), l'hypothèse (3.7) et l'inégalité  $2^{j_1} \leq n$  entraînent

$$U \leq C(2^{j_1 p/2} n^{1-p} + 2^{j_1} n^{-p/2}) \leq C 2^{j_1} n^{-p/2}.$$

Par conséquent, pour tout  $p \geq 1$ , on a

$$\sum_{k=0}^{2^{j_1}-1} \mathbb{E}_f^n(|\hat{\alpha}_{j_1,k} - \alpha_{j_1,k}|^p) \leq C 2^{j_1} n^{-p/2}.$$

Le Théorème 3.2.1 s'applique avec  $\eta = 1$  et  $j_1$  un entier défini par

$$2^{j_1} \asymp n^{1/(2s+1)}.$$

Ce qui donne la vitesse de convergence

$$\varphi_n = n^{-p/2} 2^{j_1 p/2} + 2^{-j_1 s p} \asymp n^{-\alpha_1 p}, \quad \alpha_1 = s/(2s+1).$$

Cela termine la preuve du Théorème 3.4.2. □

PREUVE DE LA PROPOSITION 3.4.1. Considérons la densité

$$g(t) = (\sigma + 1)t^\sigma, \quad \sigma > 2p^{-1}.$$

En utilisant l'égalité (4.13) et en reprenant la preuve de la Proposition 3.3.1, on montre l'existence d'une constante  $c > 0$  telle que

$$\mathcal{R}_{n,p}(B_{\pi,r}^s(L)) \geq c n^{-\alpha_3 p}, \quad \alpha_3 = (s - 1/\pi + 1/p)/(2(s - 1/\pi) + 1 + \sigma).$$

Comme  $\alpha_3 < \alpha_1 \wedge \alpha_2$  pour  $\sigma$  suffisamment grand, la preuve de la Proposition 3.4.1 est terminée. □

\* \* \*



# Étude minimax : boules de Besov pondérées

Le problème est le même que dans le chapitre 3 : on cherche à estimer une fonction inconnue perturbée par une fonction parasite. Pour ce faire, nous adoptons l'approche minimax sous le risque  $\mathbb{L}^p$  sur des boules de Besov pondérées. Nous apportons quelques compléments à l'article de Kerkycharian et Picard (2005).

## Sommaire

---

<b>4.1 Introduction . . . . .</b>	<b>60</b>
<b>4.2 Outils mathématiques . . . . .</b>	<b>61</b>
4.2.1 Classe $\mathcal{A}_p$ . . . . .	61
4.2.2 Base d'ondelettes déformées . . . . .	62
4.2.3 Boules de Besov pondérées . . . . .	64
<b>4.3 Bornes supérieures . . . . .</b>	<b>65</b>
4.3.1 Estimateur par projection en ondelettes déformées . . . .	65
4.3.2 Estimateur de seuillage dur en ondelettes déformées . . .	66
<b>4.4 Applications . . . . .</b>	<b>67</b>
4.4.1 Bruit blanc gaussien généralisé . . . . .	67
4.4.2 Régression à pas aléatoires . . . . .	69
<b>4.5 Simulations graphiques : régression à pas aléatoires . . .</b>	<b>70</b>
4.5.1 Cas où la loi des pas est connue . . . . .	70
4.5.2 Cas où la loi des pas est inconnue . . . . .	74
<b>4.6 Démonstrations . . . . .</b>	<b>77</b>
4.6.1 Preuves des résultats de la Section 4.3 . . . . .	77
4.6.2 Preuves des résultats de la Section 4.4 . . . . .	78

---

## 4.1 Introduction

### MOTIVATIONS

Lorsque qu'une fonction parasite intervient dans un modèle statistique, il est parfois difficile de calculer les vitesses minimax sous le risque  $\mathbb{L}^p$  sur des boules de Besov. Cela a été mis en relief dans le chapitre précédent avec, entre autre, le modèle de régression à pas aléatoires (3.6) pour la densité  $g : [0, 1] \rightarrow [0, \sigma + 1]$  définie par

$$g(t) = (\sigma + 1)t^\sigma, \quad \sigma > 2p^{-1}. \quad (4.1)$$

Cette absence de repère est un obstacle dans l'estimation de la fonction inconnue. Pour contourner celui-ci, plusieurs solutions sont envisageables. La plus naturelle d'entre elles consiste à reconfigurer notre approche minimax. Cette remise en cause peut aussi bien porter sur le risque comme sur l'ensemble de fonctions initialement choisi.

Ici, nous travaillons de nouveau avec le risque  $\mathbb{L}^p$  et nous nous intéressons à l'alternative proposée par les boules de Besov pondérées. La différence majeure entre celles-ci et les boules de Besov standards est qu'elles dépendent explicitement de la fonction parasite. Pour bien comprendre les raisons qui nous poussent à considérer ces ensembles de fonctions, un retour aux sources s'impose.

Plaçons nous dans le cadre du modèle de régression à pas aléatoires (3.6). Les boules de Besov pondérées ont été introduites en estimation fonctionnelle par Kerkycharian et Picard (2005). Elles ont vu le jour grâce à l'étude maxiset d'un estimateur adaptatif construit sur une base d'ondelettes déformées. Si on considère ces boules et le risque  $\mathbb{L}^p$ , l'estimateur considéré a la particularité d'atteindre une vitesse de convergence stable pour une large classe de densités  $g$ . Cette dernière est liée à une condition développée par Muckenhoupt (1972). Elle inclut les densités définies par (4.1).

Dans cette étude, nous apportons des réponses précises aux questions suivantes.

- Quelles sont les vitesses minimax sous le risque  $\mathbb{L}^p$  sur des boules de Besov pondérées ?
- Est-ce que les estimateurs construits sur une base d'ondelettes déformées sont performants en pratique ?

### PRÉSENTATION

Dans un premier temps, nous étudions les vitesses minimax sous le risque  $\mathbb{L}^p$  sur les boules de Besov pondérées. Notre objectif est de fournir des résultats applicables à certains modèles statistiques où intervient une fonction parasite. Pour amorcer notre travail, celle-ci est supposée vérifier une condition de type Muckenhoupt.

D'abord, nous évaluons les bornes supérieures grâce à l'estimateur par projection en ondelettes déformées. Puis, nous établissons les bornes inférieures en combinant un théorème de minoration élaboré par Tsybakov (2004) avec une version pondérée du théorème de Varshamov-Gilbert. Enfin, nous appliquons ces résultats au modèle de bruit blanc gaussien généralisé et au modèle de régression à pas aléatoires. Contrairement à leurs consoeurs, nous montrons que les boules de Besov pondérées fournissent des vitesses minimax stables pour une large classe de fonctions parasites. Par la même occasion, nous concluons que l'estimateur de seuillage en ondelettes déformées proposé par Kerkycharian et Picard (2005) est presque optimal pour de nombreux modèles statistiques.

Dans un deuxième temps, notre attention se focalise sur le modèle de régression à pas aléatoires. Nous présentons quelques simulations graphiques des deux estimateurs adaptatifs en ondelettes déformées introduits dans l'article de Kerkycharian et Picard (2005). L'un est explicitement construit avec la densité  $g$ . L'autre est entièrement déterminé par les observations sans connaissance a priori de  $g$ . Cela est mis en oeuvre grâce au logiciel Matlab 7 enrichi de la librairie Wavelab<sup>3</sup>. Ces simulations sont le fruit d'un travail commun avec Thomas Willer<sup>4</sup>.

## 4.2 Outils mathématiques

Le but de cette section est de poser les bases mathématiques du chapitre. Dans un premier temps, on décrit la classe  $\mathcal{A}_p$ . Dans un deuxième temps, on définit la base d'ondelettes déformées adaptée à l'intervalle  $[0, 1]$ . Enfin, on présente les boules de Besov pondérées.

### 4.2.1 Classe $\mathcal{A}_p$

DÉFINITION (POIDS)

*On appelle poids toute fonction mesurable positive sur  $[0, 1]$ .*

DÉFINITION (CLASSE  $\mathcal{A}_p$ )

*Soient  $p \in [1, \infty[$  et  $q$  tel que  $1/p + 1/q = 1$ . Un poids  $m$  appartient à la classe  $\mathcal{A}_p$  si et seulement si il existe une constante  $C > 0$  telle que, pour toute fonction mesurable  $h$  et tout sous-intervalle  $I$  de  $[0, 1]$ , on a*

$$(|I|^{-1} \int_I |h(x)| dx) \leq C (m(I)^{-1} \int_I |h(x)|^p m(x) dx)^{1/p}, \quad (4.2)$$

---

<sup>3</sup>Pour des informations complémentaires, voir : <http://www-stat.stanford.edu/~wavelab/>

<sup>4</sup>Université Paris VII. Contact e-mail : willer@math.jussieu.fr

où  $|I|$  la longueur de  $I$  et  $m$  est l'opérateur défini par  $m(I) = \int_I m(x)dx$ .

La classe  $\mathcal{A}_p$  a été introduite par Muckenhoupt (1972). Elle a connu un développement significatif grâce aux travaux de Coifman et Fefferman (1974). Mathématiquement parlant, elle caractérise la continuité de certains opérateurs intégrales de  $\mathbb{L}_m^p$  sur  $\mathbb{L}_m^p$  où, pour tout poids  $m$ , l'espace  $\mathbb{L}_m^p$  est l'ensemble des fonctions  $h : [0, 1] \rightarrow \mathbb{R}$  telles que

$$\int_0^1 |h(x)|^p m(x) dx < \infty.$$

Pour une théorie complète, voir le livre de Stein (1993).

#### 4.2.1.1 Exemples et Propriétés.

Le poids  $m : [0, 1] \rightarrow [0, \infty]$  défini par

$$m(x) = x^a, \quad a \in ]-1, p-1[,$$

appartient à la classe  $\mathcal{A}_p$ .

Si  $m \in \mathcal{A}_p$  et si  $\pi \geq p$  alors  $m \in \mathcal{A}_\pi$ .

PROPRIÉTÉ (DOUBLING)

Si  $m \in \mathcal{A}_p$  alors il existe une constante  $C > 0$  telle que, pour tout sous-intervalle  $S \subseteq B \subseteq [0, 1]$ , on a

$$m(B) \leq C|B|^p|S|^{-p}m(S). \quad (4.3)$$

Cette inégalité va jouer un rôle important dans les preuves à venir.

## 4.2.2 Base d'ondelettes déformées

D'abord, on définit la base d'ondelettes déformées. Puis, on étudie la décomposition d'une fonction sur cette base. Ensuite, on définit le poids  $w$ . Grâce à celui-ci, on traite des propriétés géométriques de ces bases en norme  $\mathbb{L}^p$ .

### 4.2.2.1 Définition

DÉFINITION (BASE D'ONDELETES DÉFORMÉES SUR L'INTERVALLE UNITÉ)

Considérons la base d'ondelettes  $\zeta$  décrite par (2.1) et  $T : [0, 1] \rightarrow [0, 1]$  une fonction mesurable croissante, bijective, absolument continue avec  $T(0) = 0$  et  $T(1) = 1$ . Pour tout entier  $j_1 \geq \tau$ , on définit la base d'ondelettes déformées sur l'intervalle  $[0, 1]$  par

$$\zeta^T = \{\phi_{j_1, k}(T(x)), k = 0, \dots, 2^{j_1} - 1; \psi_{j, k}(T(x)), j = j_1, \dots, \infty, k = 0, \dots, 2^j - 1\}.$$

Précisons que si  $T$  n'est pas la fonction identité alors la base  $\zeta^T$  n'est pas orthonormale. Dorénavant,  $T$  fait référence à la fonction précédemment définie.

#### 4.2.2.2 Poids $w$

DÉFINITION (POIDS  $w$ )

On définit le poids  $w$  par

$$w(x) = 1/(T'(T^{-1}(x))), \quad x \in [0, 1], \quad (4.4)$$

où  $T'$  désigne la dérivée de  $T$  et  $T^{-1}$  désigne la fonction inverse.

Le poids  $w$  fait le lien entre la classe  $\mathcal{A}_p$  et certaines propriétés géométriques de la base d'ondelettes déformées  $\zeta^T$  en norme  $\mathbb{L}^p$ . Une équivalence à la condition  $w \in \mathcal{A}_p$  est donnée ci-dessous.

PROPRIÉTÉ (EQUIVALENCE À LA CONDITION  $w \in \mathcal{A}_p$ )

Soient  $p \in ]1, \infty[$  et  $q$  tel que  $1/p + 1/q = 1$ . On a  $w \in \mathcal{A}_p$  si et seulement si il existe une constante  $C > 0$  telle que, pour tout sous-intervalle  $I$  de  $[0, 1]$ , on a

$$(|I|^{-1} \int_I T'(x)^q dx)^{1/q} \leq C |I|^{-1} \int_I T'(x) dx.$$

Autrement dit, si  $w \in \mathcal{A}_p$ , alors  $T'$  vérifie une inégalité de Hölder inversée. Précisons que la condition  $w \in \mathcal{A}_p$  sera récurrente dans notre étude. Dorénavant,  $w$  fait référence au poids défini par (4.4).

#### 4.2.2.3 Décomposition

Supposons que  $w \in \mathcal{A}_p$ . Pour tout entier  $j_1 \geq \tau$ , une fonction  $f$  de  $\mathbb{L}^2([0, 1])$  se décompose sur la base  $\zeta^T$  comme

$$f(x) = \sum_{k=0}^{2^{j_1}-1} \alpha_{j_1,k}^T \phi_{j_1,k}(T(x)) + \sum_{j=j_1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k}^T \psi_{j,k}(T(x)),$$

où

$$\alpha_{j,k}^T = \int_0^1 f(T^{-1}(x)) \phi_{j,k}(x) dx \quad \text{et} \quad \beta_{j,k}^T = \int_0^1 f(T^{-1}(x)) \psi_{j,k}(x) dx.$$

#### 4.2.2.4 Propriétés géométriques

Quelques propriétés géométriques satisfaites par la base  $\zeta^T$  sont décrites ci-dessous.

PROPRIÉTÉ (INCONDITIONNALITÉ ; TEMLYAKOV)

Soit  $p \in ]1, \infty[$ . Supposons que  $w \in \mathcal{A}_p$ . Alors la base  $\zeta^T$  vérifie la propriété d'inconditionnalité et la propriété de Temlyakov. Autrement dit, les inégalités (2.3) et (2.4) sont vérifiées si on remplace  $\psi_{j,k}$  par  $\psi_{j,k}(T(\cdot))$  et  $\phi_{j,k}$  par  $\phi_{j,k}(T(\cdot))$ .

Le premier point a été montré par Garcia-Cuerva et Martell (2001) et le second point par Kerkycharian et Picard (2003).



**PROPRIÉTÉ (ÉLÉMENTAIRE)**

Soit  $p \in ]1, \infty[$ . Supposons que  $w \in \mathcal{A}_p$ . Soient  $\theta^T \in \{\alpha^T, \beta^T\}$  et  $h \in \{\phi, \psi\}$ . Alors, pour tout entier  $j \geq \tau$ , on a

$$\left\| \sum_{k=0}^{2^j-1} \theta_{j,k}^T h_{j,k}(T(\cdot)) \right\|_p^p \asymp 2^{jp/2} \sum_{k=0}^{2^j-1} |\theta_{j,k}^T|^p w(I_{j,k}), \quad (4.5)$$

où  $w(I_{j,k}) = \int_{I_{j,k}} w(x) dx$  et  $I_{j,k} = [k2^{-j}, (k+1)2^{-j}]$ .

Pour une preuve détaillée, voir de nouveau l'article de Kerkycharian et Picard (2003). Dorénavant,  $I_{j,k}$  désignera cet intervalle.

### 4.2.3 Boules de Besov pondérées

Cette sous-section présente les principaux ensembles de fonctions étudiés : les boules de Besov pondérées. Sous certaines hypothèses sur  $T$ , celles-ci se caractérisent simplement à l'aide de la base  $\zeta^T$ .

#### 4.2.3.1 Boules de Besov pondérées et module de continuité

**DÉFINITION (BOULES DE BESOV PONDÉRÉES)**

Soient  $L \in ]0, \infty[$ ,  $s \in ]0, \infty[$ ,  $\pi \in [1, \infty]$  et  $u = \lfloor s \rfloor + 1$ . Pour toute fonction mesurable  $f : [0, 1] \rightarrow \mathbb{R}$ , on pose

$$\Delta_h(f)(x) = f(x+h) - f(x) = (\mathcal{T}_h - I)(f)(x),$$

où  $\mathcal{T}_h$  désigne l'opérateur de translation de pas  $h$ . On pose

$$\Delta_h^u(f)(x) = (\mathcal{T}_h - I)^u(f)(x) = \sum_{k=0}^u \binom{u}{k} (-1)^{u-k} f(x+kh).$$

On définit le  $u$ -ième module de continuité pondéré par

$$\rho^u(t, f, T, \pi) = \sup_{|h| \leq t} \left( \int_{J_{u,h}^*} |\Delta_h^u(f(T^{-1}))(x)|^\pi w(x) dx \right)^{1/\pi},$$

où  $J_{u,h}^* = \{x \in [0, 1] \mid T(x) + uh \in [0, 1]\}$ . Une fonction  $f$  appartient à la boule de Besov pondérée  $B_{\pi,r}^{s,T}(L)$  si et seulement si

$$\|f\|_\pi + \begin{cases} \left( \int_0^1 (t^{-s} \rho^u(t, f, T, \pi))^r t^{-1} dt \right)^{1/r} \leq L, & \text{si } r \in [1, \infty[, \\ \sup_{t \in [0,1]} t^{-s} \rho^u(t, f, T, \pi) \leq L, & \text{si } r = \infty. \end{cases}$$

Les boules de Besov pondérées ont été introduites en Analyse par Qui (1982) et furent récemment utilisées dans un cadre statistique par Kerkycharian et Picard (2005).

**PROPRIÉTÉ (INCLUSION)**

Soient  $p \in ]1, \infty[$ ,  $\pi \in [1, \infty]$ ,  $s \in ]1/\pi, \infty]$  et  $r \in [1, \infty]$ . Si  $w \in \mathcal{A}_p$  alors on a l'inclusion

$$B_{\pi,r}^{s,T}(L) \subseteq B_{p,\infty}^{s,T}(L).$$

### 4.2.3.2 Ondelettes déformées et boules de Besov pondérées

Les boules de Besov pondérées ont l'avantage de s'exprimer simplement à l'aide de la base d'ondelettes déformées  $\zeta^T$ .

PROPRIÉTÉ (CARACTÉRISATION EN ONDELETTES DÉFORMÉES)

Soit  $\pi \in ]1, \infty[$ . Supposons que  $w$  vérifie la condition  $\mathcal{A}_\pi$ . Une fonction  $f$  appartient à la boule de Besov pondérée  $B_{\pi,r}^{s,T}(L)$  si et seulement si les coefficients en ondelettes déformées associés vérifient

$$\begin{aligned} & \left( \sum_{k=0}^{2^\tau-1} |\alpha_{\tau,k}^T|^\pi w(I_{\tau,k}) \right)^{1/\pi} \\ & + \begin{cases} \left( \sum_{j=\tau}^{\infty} (2^{j(s+1/2)} (\sum_{k=0}^{2^j-1} |\beta_{j,k}^T|^\pi w(I_{j,k}))^{1/\pi})^r \right)^{1/r} \leq L_*, & \text{si } r \in [1, \infty], \\ \sup_{j=\tau}^{\infty} 2^{j(s+1/2)} (\sum_{k=0}^{2^j-1} |\beta_{j,k}^T|^\pi w(I_{j,k}))^{1/\pi} \leq L_*, & \text{si } r = \infty, \end{cases} \end{aligned}$$

où  $L_* > 0$  est une constante proportionnelle à  $L$ . Cette équivalence est vraie pour  $s \in ]q(w), N[$  où  $q(w)$  est le plus petit  $v > 1$  tel que  $w$  vérifie la condition  $\mathcal{A}_v$  si  $w$  n'est pas l'identité et 0 sinon.

Voir l'article de Kerkyacharian et Picard (2005).

## 4.3 Bornes supérieures

On considère un modèle général à partir duquel on peut estimer les coefficients d'ondelettes déformées  $\alpha_{j,k}^T$  et  $\beta_{j,k}^T$  de la fonction inconnue  $f : [0, 1] \rightarrow \mathbb{R}$ . Les estimateurs correspondants sont notés  $\hat{\alpha}_{j,k}^T$  et  $\hat{\beta}_{j,k}^T$ . La fonction  $T$  peut dépendre de certaines données du modèle.

### 4.3.1 Estimateur par projection en ondelettes déformées

DÉFINITION (ESTIMATEUR PAR PROJECTION EN ONDELETTES DÉFORMÉES)

Soit  $j_1 \in \{\tau, \dots, \infty\}$ . On définit l'estimateur par projection en ondelettes déformées  $\hat{f}_n^{l,T} : [0, 1] \rightarrow \mathbb{R}$  par

$$\hat{f}_n^{l,T}(x) = \sum_{k=0}^{2^{j_1}-1} \hat{\alpha}_{j_1,k}^T \phi_{j_1,k}(T(x)). \quad (4.6)$$

Le Théorème 4.3.1 ci-dessous étudie les bornes supérieures de l'estimateur par projection en ondelettes déformées sous le risque  $\mathbb{L}^p$  sur  $B_{\pi,r}^{s,T}(L)$  avec  $\pi \geq p$  où  $T$  désigne une fonction convenablement choisie.

### THÉORÈME 4.3.1

Soit  $p \in ]1, \infty[$ . Plaçons nous dans le cadre statistique général décrit au début de la présente section. Supposons que  $w \in \mathcal{A}_p$ , que  $f \in B_{\pi,r}^{s,T}(L)$  et que l'estimateur  $\hat{\alpha}_{j_s,k}^T$  vérifie

**HYPOTHÈSE (D)** Il existe une constante  $C > 0$  telle que, pour  $k \in \{0, \dots, 2^{j_s} - 1\}$  et  $n$  suffisamment grand, on a

$$\mathbb{E}_f^n(|\hat{\alpha}_{j_s,k}^T - \alpha_{j_s,k}^T|^p) \leq Cn^{-p/2},$$

où  $j_s$  est un entier tel que  $2^{j_s} \asymp n^{1/(2s+1)}$ .

Considérons l'estimateur  $\hat{f}_n^{l,T}$  défini par (4.6). Alors il existe une constante  $C > 0$  telle que, pour tout  $\pi \in [1, \infty]$ ,  $s \in [q(w), \infty]$ ,  $r \in [1, \infty]$  et  $n$  suffisamment grand, on a

$$\sup_{f \in B_{\pi,r}^{s,T}(L)} \mathbb{E}_f^n(\|\hat{f}_n^{l,T} - f\|_p^p) \leq Cn^{-\alpha_1 p},$$

où  $\alpha_1 = s/(2s+1)$ .

La vitesse de convergence présentée dans le Théorème 4.3.1 est minimax pour de nombreux modèles statistiques. En particulier, avec un choix convenable de  $T$ , elle est minimax pour le modèle de régression à pas aléatoires et le modèle de bruit blanc gaussien généralisé. Cela sera abordé en détail dans les sections suivantes.

Notons que si  $T$  est la fonction identité alors le Théorème 4.3.1 délivre un résultat minimax standard.

## 4.3.2 Estimateur de seuillage dur en ondelettes déformées

Cette sous-section rappelle quelques résultats minimax et maxisets liés à l'estimateur de seuillage dur en ondelettes déformées développé par Kerkycharian et Picard (2005).

### 4.3.2.1 Construction

**DÉFINITION (ESTIMATEUR DE SEUILLAGE DUR EN ONDELETES DÉFORMÉES)**

Soient  $\nu \in ]0, 1]$  et  $\mu$  un réel positif. Considérons des entiers  $j_1$  et  $j_2$  tels que

$$j_1 = \tau \quad \text{et} \quad 2^{j_2} \asymp (n/\log n)^\nu.$$

On définit l'estimateur de seuillage dur en ondelettes déformées  $\hat{f}_n^{h,T} : [0, 1] \rightarrow \mathbb{R}$  par

$$\hat{f}_n^{h,T}(x) = \sum_{k=0}^{2^{j_1}-1} \hat{\alpha}_{j_1,k}^T \phi_{j_1,k}(T(x)) + \sum_{j=j_1}^{j_2} \sum_{k=0}^{2^j-1} \hat{\beta}_{j,k}^T 1_{\{|\hat{\beta}_{j,k}^T| \geq \mu \sqrt{(\log n/n)}\}} \psi_{j,k}(T(x)). \quad (4.7)$$

Tous les paramètres mis en jeu sont indépendants de la régularité supposée de  $f$ . L'estimateur  $\hat{f}_n^{h,T}$  est donc adaptatif.

### 4.3.2.2 Résultats minimax et maxiset

Si on remplace  $\hat{\alpha}_{j,k}$  par  $\hat{\alpha}_{j,k}^T$ ,  $\hat{\beta}_{j,k}$  par  $\beta_{j,k}^T$ ,  $\psi_{j,k}(\cdot)$  par  $\psi_{j,k}(T(\cdot))$ ,  $\phi_{j,k}(\cdot)$  par  $\phi_{j,k}(T(\cdot))$  et a fortiori  $\hat{f}_n^h$  par  $\hat{f}_n^{h,T}$  et  $B_{\pi,r}^s(L)$  par  $B_{\pi,r}^{s,T}(L)$  alors, sous

- l’hypothèse que  $w \in \mathcal{A}_p$ ,
- des hypothèses analogues à (01) et (02) de la sous-section 2.4.2,

on est en mesure de poser des résultats analogues aux Théorèmes 2.4.1 pour  $\pi \geq p$  et 2.4.2. Voir l’article de Kerkyacharian et Picard (2005).

## 4.4 Applications

Cette section est consacrée au modèle de bruit blanc gaussien généralisé et au modèle de régression à pas aléatoires.

### 4.4.1 Bruit blanc gaussien généralisé

#### 4.4.1.1 Vitesse minimax

Dans le cadre du modèle de bruit blanc gaussien généralisé défini par (3.3), le Théorème 4.4.1 ci-dessous isole la vitesse minimax sous le risque  $\mathbb{L}^p$  sur  $B_{\pi,r}^{s,T}(L)$  avec  $\pi \geq p$  où  $T$  désigne une fonction convenablement choisie.

#### THÉORÈME 4.4.1

Soit  $p \in ]1, \infty[$ . Considérons le modèle de bruit blanc gaussien généralisé défini par (3.3). Supposons que  $\int_0^1 v(x)^{-2} dx = 1$ . On définit la fonction  $H : [0, 1] \rightarrow [0, 1]$  par

$$H(t) = \int_0^t v(x)^{-2} dx.$$

On suppose qu’elle est bijective et que

$$v^2(H^{-1}(\cdot)) \in \mathcal{A}_p. \quad (4.8)$$

Alors il existe deux constantes  $C > 0$  et  $c > 0$  telles que, pour tout  $\pi \in [p, \infty]$ ,  $s \in [q(w), \infty]$ ,  $r \in [1, \infty]$  et  $n$  suffisamment grand, on a

$$cn^{-\alpha_1 p} \leq \mathcal{R}_{n,p}(B_{\pi,r}^{s,H}(L)) \leq Cn^{-\alpha_1 p},$$

où  $\alpha_1 = s/(2s + 1)$ .

Les bornes supérieures sont calculées grâce au Théorème 4.3.1. La borne inférieure résulte du Théorème 3.2.3 combiné avec une version pondérée du Théorème de Vashamov-Gilbert.

On constate que la vitesse minimax obtenue est stable pour une large classe de fonctions  $v$ . En rapport avec le chapitre précédent, celle-ci comprend de nombreuses fonctions vérifiant  $v \notin \mathbb{L}^{p_*}([0, 1])$  avec  $p_* = p \vee 2$ . Un exemple précis est étudié à la sous-sous-section 4.4.1.3. On interprète ce résultat de la façon suivante : si  $f \in B_{\pi, r}^{s, H}(L)$  alors elle ne varie pas brutalement dans le voisinage où  $v$  tend vers l'infini en un point de l'intervalle  $[0, 1]$ .

#### 4.4.1.2 Résultat adaptatif

Le Théorème 4.4.2 ci-dessous détermine la borne supérieure atteinte par l'estimateur de seuillage dur en ondelettes déformées sous le risque  $\mathbb{L}^p$  sur  $B_{\pi, r}^{s, H}(L)$  avec  $\pi \geq p$ . La fonction  $H$  est définie comme dans le Théorème 4.4.1.

##### THÉORÈME 4.4.2

*Soit  $p \in ]1, \infty[$ . Considérons le modèle de bruit blanc gaussien généralisé défini par (3.3). Adoptons les hypothèses et les notations du Théorème 4.4.1. Posons  $\nu = 1$ ,*

$$\hat{\alpha}_{j,k} = \int_0^1 \phi_{j,k}(H(t))v(t)^{-1}dY(t) \quad \text{et} \quad \hat{\beta}_{j,k} = \int_0^1 \psi_{j,k}(H(t))v(t)^{-1}dY(t). \quad (4.9)$$

*Alors il existe une constante  $C > 0$  telle que, pour tout  $\pi \in [p, \infty]$ ,  $s \in [q(w), \infty[$ ,  $r \in [1, \infty]$  et  $n$  suffisamment grand, on a*

$$\sup_{f \in B_{\pi, r}^{s, H}(L)} \mathbb{E}_f^n(\|\hat{f}_n^{h, H} - f\|_p^p) \leq C(\log n/n)^{\alpha_1 p},$$

où  $\alpha_1 = s/(2s + 1)$ .

Le Théorème 4.4.2 montre que si  $f$  appartient aux boules de Besov pondérées alors on peut construire un estimateur adaptatif presque optimal pour une large classe de fonctions  $v$ .

#### 4.4.1.3 Exemple

Pour illustrer cette largesse, étudions une famille de fonctions qui posait problème pour le calcul des vitesses minimax sur des boules de Besov standards.

Soit  $p \in ]1, \infty[$ . Considérons le modèle (3.3) avec la fonction  $v$  définie par

$$v(t) = (\sigma + 1)^{-1/2}t^{-\sigma/2}, \quad \sigma > 1/p - 1. \quad (4.10)$$

On a

$$H(t) = t^{\sigma+1}, \quad H^{-1}(t) = t^{1/(\sigma+1)} \quad \text{et} \quad w(t) = v^2(H^{-1}(t)) = (\sigma + 1)^{-1}t^{-\sigma/(\sigma+1)}.$$

Puisque  $-\sigma/(\sigma + 1) \in ]-1, p - 1[$ , on a  $w \in \mathcal{A}_p$ . On est donc en mesure d'appliquer les Théorèmes 4.4.1 et 4.4.2.

TAB. 4.1 – Vitesses minimax obtenues sur les boules de Besov standards et pondérées pour  $\pi = p$  avec  $p > 2$ . On considère le modèle de bruit blanc gaussien généralisé défini avec la fonction (4.10).

Modèle (3.3) avec $v(t) = (\sigma + 1)^{-1/2}t^{-\sigma/2}$	$\mathcal{R}_{n,p}(\mathcal{E}) \asymp n^{-\alpha p}$	
	$\mathcal{E}$	$\alpha$
$\sigma \in ]0, 2p^{-1}[$	$B_{p,r}^s(L)$	$s/(2s + 1)$
$\sigma \in ]2p^{-1}, 1[$		$s/(2s + 1 + \sigma - 2/p)$
$\sigma \in ]1, \infty[$		Indéterminé
$\sigma \in ]0, \infty[$	$B_{p,r}^{s,H}(L)$	$s/(2s + 1)$

#### 4.4.2 Régression à pas aléatoires

Dans le cadre du modèle de régression à pas aléatoires défini par (3.6), le Théorème 4.4.3 ci-dessous isole la vitesse minimax sous le risque  $\mathbb{L}^p$  sur  $B_{\pi,r}^{s,T}(L)$  avec  $\pi \geq p$  où  $T$  désigne une fonction convenablement choisie.

##### THÉORÈME 4.4.3

Soit  $p \in ]1, \infty[$ . Considérons le modèle de régression à pas aléatoires défini par (3.6). On définit la fonction  $G : [0, 1] \rightarrow [0, 1]$  par

$$G(t) = \mathbb{P}_f^n(X_1 \leq t) = \int_0^t g(x)dx.$$

Supposons qu'elle soit bijective et que

$$1/g(G^{-1}(\cdot)) \in \mathcal{A}_p. \quad (4.11)$$

Alors il existe deux constantes  $C > 0$  et  $c > 0$  telles que, pour tout  $\pi \in [p, \infty]$ ,  $s \in [q(w), \infty[$ ,  $r \in [1, \infty]$  et  $n$  suffisamment grand, on a

$$cn^{-\alpha_1 p} \leq \mathcal{R}_{n,p}(B_{\pi,r}^{s,G}(L)) \leq Cn^{-\alpha_1 p},$$

où  $\alpha_1 = s/(2s + 1)$ .

Les bornes supérieures sont calculées grâce au Théorème 4.3.1. La borne inférieure résulte du Théorème 3.2.3 combiné avec une version pondérée du Théorème de Varshamov-Gilbert.

On interprète ce résultat de la façon suivante : si  $f \in B_{\pi,r}^{s,G}(L)$  alors elle ne varie pas brutalement dans le voisinage où  $g$  s'écrase fortement sur l'axe des abscisses.

REMARQUE : Dans le cadre statistique ainsi posé, on prouve que les estimateurs de seuillage dur en ondelettes déformées construits par Kerkycharian et Picard (2005) sont presque optimaux.

En procédant comme dans l'exemple de la sous-section précédente, on montre que la densité  $g : [0, 1] \rightarrow [0, \sigma + 1]$  définie par

$$g(t) = (\sigma + 1)t^\sigma, \quad \sigma > 1/p - 1,$$

vérifie la condition (4.11). On peut donc appliquer le Théorème 4.4.3. Ainsi, grâce aux boules de Besov pondérées, on obtient des vitesses minimax stables pour une large classe de densités contrairement à l'approche minimax sur des boules de Besov standards. Voir le chapitre précédent.

TRANSITION : Ce qui précède fournit quelques apports théoriques à l'article de Kerkycharian et Picard (2005). L'aspect pratique de certains estimateurs de seuillage dur en ondelettes déformées est étudié ci-après.

## 4.5 Simulations graphiques : régression à pas aléatoires

Dans cette section, on considère le modèle de régression à pas aléatoires défini par

$$Y_i = f(X_i) + s_* z_i, \quad i = 1, \dots, n,$$

où la quantité  $s_* = 0,04$  désigne la valeur du niveau de bruit. Les autres paramètres sont décrits comme dans (3.6).

Nous proposons plusieurs simulations graphiques de deux estimateurs de seuillage dur en ondelettes déformées développés par Kerkycharian et Picard (2005). Le premier dépend explicitement de  $g$  alors que le second, plus réaliste, est entièrement construit à partir des données.

### 4.5.1 Cas où la loi des pas est connue

#### 4.5.1.1 Estimateur

On considère l'estimateur de seuillage dur en ondelettes déformées défini par (4.7). On prend  $\nu = 2^{-1}$ ,

$$\hat{\alpha}_{j,k}^G = n^{-1} \sum_{i=1}^n Y_i \phi_{j,k}(G(X_i)) \quad \text{et} \quad \hat{\beta}_{j,k}^G = n^{-1} \sum_{i=1}^n Y_i \psi_{j,k}(G(X_i)),$$

où  $G$  est la fonction de répartition de  $X_1$ .

On pose  $\mu = \hat{s}\sqrt{2}$  où  $\hat{s}$  désigne l'estimateur de Rice défini par

$$\hat{s} = ((2n - 2)^{-1} \sum_{i=2}^n (Y_{(i)} - Y_{(i-1)})^2)^{1/2}.$$

Chaque  $Y_{(i)}$  désigne la valeur de  $(Y_1, \dots, Y_n)$  associée à la  $i$  ième plus grande coordonnée du vecteur  $(X_1, \dots, X_n)$ .

#### 4.5.1.2 Configuration

On prend  $n = 2^{10}$ . La base d'ondelettes considérée est la *Symlet*[8]. C'est celle qui nous donne les meilleurs résultats.

Les variables aléatoires  $(X_1, \dots, X_n)$  sont simulées via la méthode du rejet.

Les fonctions  $f$  considérées sont Doppler, Bumps, Heavisine, Wave et DoubleSpikes. Les trois premières sont standards. Les deux dernières sont définies comme suit :

- *Wave*.  $f(x) = 2^{-1} + 0,2 \cos(4\pi x) + 0,1 \cos(24\pi x)$ ,
- *DoubleSpikes*.  $f(x) = (2x + 2^{-1})1_{\{0 \leq x \leq 0,15\}} + (-12(x - 0,15) + 0,8)1_{\{0,15 \leq x \leq 0,2\}} + 0,21_{\{0,2 \leq x \leq 0,5\}} + (6(x - 0,5) + 0,2)1_{\{0,5 \leq x \leq 0,6\}} + (-10(x - 0,6) + 0,8)1_{\{0,6 \leq x \leq 0,65\}} + (-2^{-1}(x - 0,65) + 0,3)1_{\{0,65 \leq x \leq 0,85\}} + (2(x - 0,85) + 0,2)1_{\{0,85 \leq x \leq 1\}}$ .

Dans la légende de nos tableaux, la constante  $c$  vérifie  $\int_0^1 g(t)dt = 1$ .



### 4.5.1.3 Graphiques

FIG. 4.1 – La fonction initiale est Doppler et la densité est  $g(x) = 1_{[0,1]}(x)$ .

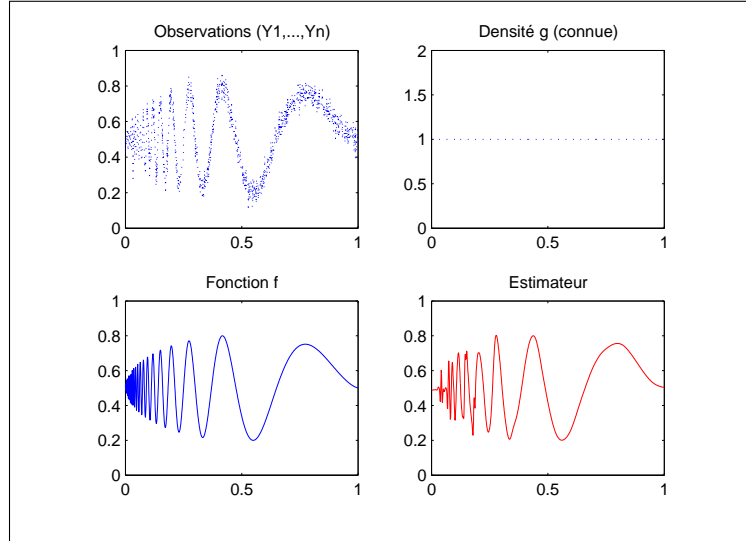


FIG. 4.2 – La fonction initiale est Wave et la densité est  $g(x) = c|x - 2^{-1}|^3 1_{[0,1]}(x)$ .

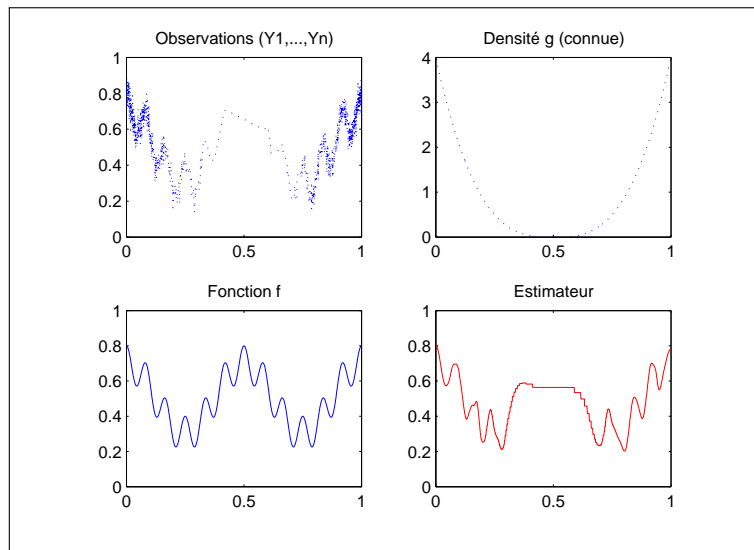


FIG. 4.3 – La fonction initiale est Heavisine et la densité est  $g(x) = c|x - 2^{-1}|1_{[0,1]}(x)$

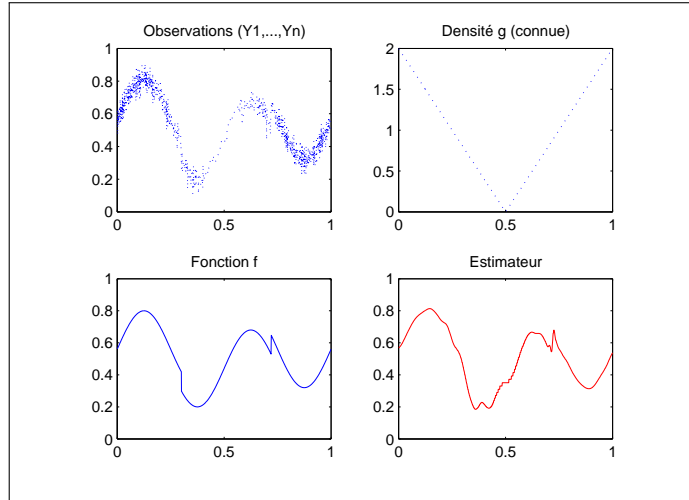
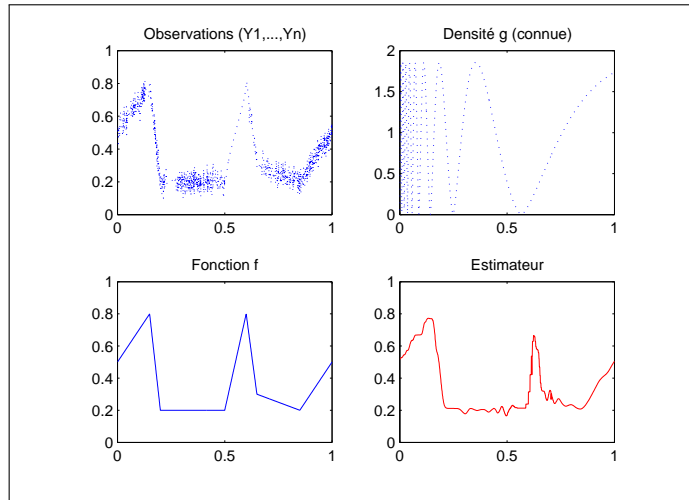


FIG. 4.4 – La fonction initiale est DoubleSpikes et la densité est  $g(x) = c|(\sin(2/(x + 0,07)))|^{3/2}1_{[0,1]}(x)$



Les Figures (4.1), (4.2), (4.3) et (4.4) illustrent les bonnes performances visuelles de l'estimateur de seuillage dur en ondelettes déformées. La première se place dans le cas où les pas suivent une loi uniforme. Dans les trois autres, on prend des densités  $g$  qui s'écrasent plus ou moins fortement sur l'axe des abscisses. On constate que la reconstruction est assez fidèle excepté dans les zones où  $g$  s'écrase. Comme les observations y sont en nombre réduit, ce manque de précision est tout à fait normal.

## 4.5.2 Cas où la loi des pas est inconnue

Ici, on se place dans une situation plus réaliste que la précédente. On suppose que  $g$  est inconnue, ce qui est généralement le cas en pratique.

### 4.5.2.1 Estimateur

DÉFINITION (ESTIMATEUR DE SEUILLAGE DUR 'DATA DRIVEN')

Soient  $j_1$  et  $j_2$  des entiers vérifiant

$$j_1 = \tau \quad \text{et} \quad 2^{j_2} \asymp (n/\log n)^{1/2}.$$

On définit l'estimateur de seuillage dur "data driven"  $\hat{f}_n^{h,\hat{G}} : [0, 1] \rightarrow [0, 1]$  par

$$\hat{f}_n^{h,\hat{G}}(x) = \sum_{k=0}^{2^{j_1}-1} \hat{\alpha}_{j_1,k}^{\hat{G}} \phi_{j_1,k}(\hat{G}(x)) + \sum_{j=j_1}^{j_2} \sum_{k=0}^{2^j-1} \hat{\beta}_{j,k}^{\hat{G}} 1_{\{|\hat{\beta}_{j,k}^{\hat{G}}| \geq \mu \sqrt{(\log n/n)}\}} \psi_{j,k}(\hat{G}(x)),$$

où

$$\hat{\alpha}_{j,k}^{\hat{G}} = n^{-1} \sum_{i=1}^n Y_i \phi_{j,k}(\hat{G}(X_i)), \quad \text{et} \quad \hat{\beta}_{j,k}^{\hat{G}} = n^{-1} \sum_{i=1}^n Y_i \psi_{j,k}(\hat{G}(X_i)).$$

$\hat{G} : [0, 1] \rightarrow [0, 1]$  est la fonction de répartition empirique définie par

$$\hat{G}(t) = n^{-1} \sum_{i=1}^n 1_{\{X_i \leq t\}}.$$

On reprend la même configuration de la sous-section précédente. L'estimateur considéré sera noté Estimateur\*.

L'estimateur précédent n'est pas tout à fait celui proposé par Kerkycharian et Picard (2005) mais le principe de construction est rigoureusement le même.

### 4.5.2.2 Graphiques

FIG. 4.5 – La fonction initiale est DoubleSpikes et la densité est  $g(x) = 1_{[0,1]}(x)$ .

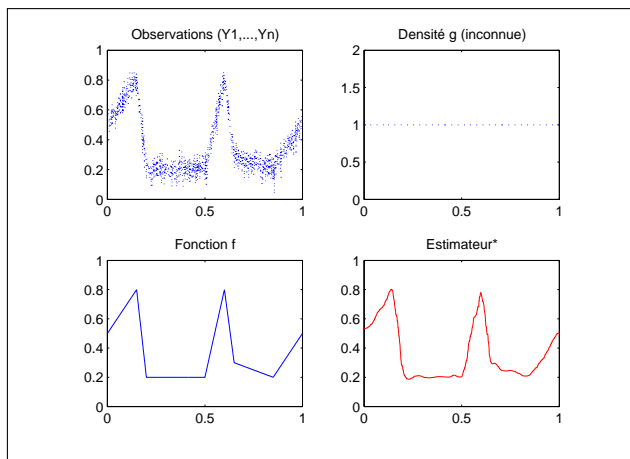


FIG. 4.6 – La fonction initiale est Doppler et la densité est  $g(x) = c(1 + 0.2 \sin(4\pi x) + 0,1 \text{irreg}(\max(1, \text{round}(100x))))1_{[0,1]}(x)$ . Le dernier terme étant issu d'une macro Matlab.

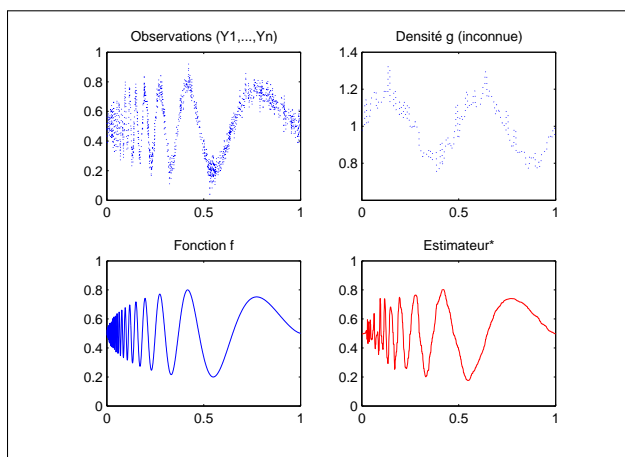


FIG. 4.7 – La fonction initiale est un dérivé de Bumps et la densité est  $g(x) = c|\sin(2/(x + 0,07))|^{3/2}1_{[0,1]}(x)$ .

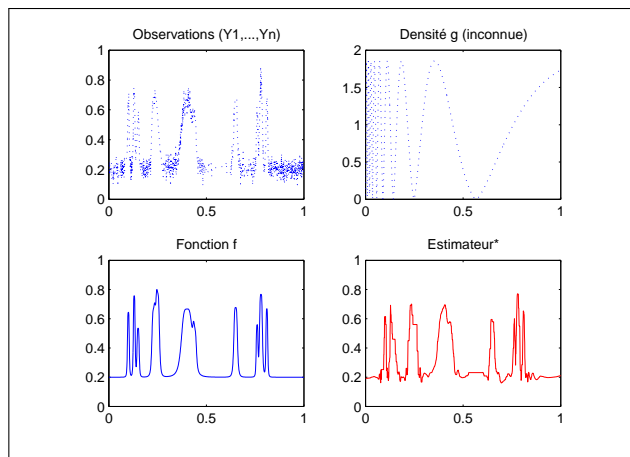
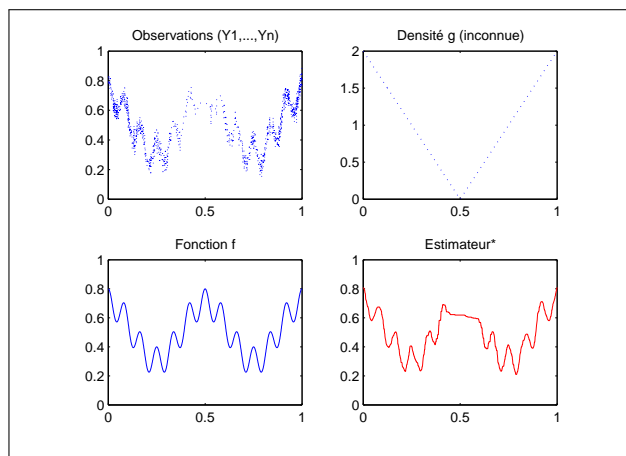


FIG. 4.8 – La fonction initiale est Wave et la densité est  $g(x) = c|x - 2^{-1}|1_{[0,1]}(x)$ .



Dans les Figures (4.5), (4.6), (4.7) et (4.8), on constate le bon rendu visuel de l'estimateur de seuillage dur data driven. Précisons que cette construction est simple à implémenter et que les temps de calculs sont rapides.

En guise de perspective, il serait intéressant de comparer les performances numériques de cet estimateur avec les estimateurs par polynômes locaux élaborés par Gaïffas (2005) et ceux reposant sur une base d'ondelettes adaptée aux  $(X_1, \dots, X_n)$  développés par Delouille et al. (2001) et Delouille et al. (2004).

## 4.6 Démonstrations

Les constantes  $c$  et  $C$  ne dépendent ni de  $f$ , ni de  $n$ . Elles peuvent prendre des valeurs différentes d'un terme à l'autre. On suppose que l'entier  $n$  est suffisamment grand.

### 4.6.1 Preuves des résultats de la Section 4.3

**Preuve du Théorème 4.3.1.** Supposons que  $w \in \mathcal{A}_p$ . En utilisant une inégalité élémentaire de convexité, on a

$$\mathbb{E}_f^n(\|\hat{f}_n^{l,T} - f\|_p^p) \leq 2^{p-1}(\mathbb{E}_f^n(\|\hat{f}_n^{l,T} - P_{j_s}^T(f)\|_p^p) + \|P_{j_s}^T(f) - f\|_p^p) = C(Q_1 + Q_2),$$

où  $P_{j_s}^T(f) : [0, 1] \rightarrow \mathbb{R}$  est défini par

$$P_{j_s}^T(f)(x) = \sum_{k=0}^{2^{j_s}-1} \alpha_{j_s,k}^T \phi_{j_s,k}(T(x)).$$

Analysons les majorations des termes  $Q_1$  et  $Q_2$ .

– *Majoration de  $Q_1$ .* En décomposant  $f$  sur  $\zeta$  à partir du niveau  $j_s$  et en utilisant l'inégalité (4.5), on obtient

$$Q_1 = \mathbb{E}_f^n(\|\sum_{k=0}^{2^{j_s}-1} (\hat{\alpha}_{j_s,k}^T - \alpha_{j_s,k}^T) \phi_{j_s,k}(T(\cdot))\|_p^p) \leq C 2^{j_s p/2} \sum_{k=0}^{2^{j_s}-1} \mathbb{E}_f^n(|\hat{\alpha}_{j_s,k}^T - \alpha_{j_s,k}^T|^p) w(I_{j_s,k}),$$

où  $w(I_{j,k}) = \int_{I_{j,k}} w(t) dt$ . Par l'hypothèse (D), il vient

$$\sum_{k=0}^{2^{j_s}-1} \mathbb{E}_f^n(|\hat{\alpha}_{j_s,k}^T - \alpha_{j_s,k}^T|^p) w(I_{j_s,k}) \leq C n^{-p/2} \sum_{k=0}^{2^{j_s}-1} w(I_{j_s,k}) = C n^{-p/2}.$$

D'où

$$Q_1 \leq C 2^{j_s p/2} n^{-p/2} \leq C n^{-\alpha_1 p}.$$

– *Majoration de  $Q_2$ .* L'inégalité de Minkowski et l'inclusion  $B_{\pi,r}^{s,T}(L) \subseteq B_{p,\infty}^{s,T}(L)$  entraînent

$$Q_2 \leq C \left( \sum_{j=j_s}^{\infty} 2^{j/2} \left( \sum_{k=0}^{2^j-1} |\beta_{j,k}^H|^p w(I_{j,k}) \right)^{1/p} \right)^p \leq C \left( \sum_{j=j_s}^{\infty} 2^{-js} \right)^p \leq C 2^{-j_s s p} \leq C n^{-\alpha_1 p}.$$

Par conséquent

$$\sup_{f \in B_{\pi,r}^{s,T}(L)} \mathbb{E}_f^n(\|\hat{f}_n^{l,T} - f\|_p^p) \leq C(2^{j_s p/2} n^{-p/2} + 2^{-j_s s p}) \leq C n^{-\alpha_1 p},$$

avec  $\alpha_1 = s/(2s+1)$ . Cela termine la preuve du Théorème 4.3.1.  $\square$

## 4.6.2 Preuves des résultats de la Section 4.4

**Preuve du Théorème 4.4.1.** Étudions séparément la preuve de la borne supérieure et la preuve de la borne inférieure.

• *Borne supérieure.* L'objectif est d'appliquer le Théorème 4.3.1. Montrons que l'hypothèse (D) est satisfaite. Considérons l'estimateur

$$\hat{\alpha}_{j_s,k}^H = \int_0^1 \phi_{j_s,k}(H(t)) v(t)^{-1} dY(t),$$

où  $j_s$  est un entier tel que  $2^{j_s} \asymp n^{1/(2s+1)}$ .

Par le changement de variable  $y = H^{-1}(t)$ , il vient

$$\hat{\alpha}_{j_s,k}^H - \alpha_{j_s,k}^H = n^{-1/2} \int_0^1 \phi_{j_s,k}(H(t)) v(t)^{-1} dW(t) \sim \mathcal{N}(0, n^{-1}).$$

Par une inégalité gaussienne élémentaire, on obtient

$$\mathbb{E}_f^n(|\hat{\alpha}_{j_s,k}^H - \alpha_{j_s,k}^H|^p) \leq C n^{-p/2}.$$

On conclut en utilisant le Théorème 4.3.1.

• *Borne inférieure.* Soit  $\pi \geq p$ . L'objectif est d'appliquer le Théorème 3.2.3.

– *Hypothèse (C1).* Soit  $j_3$  un entier tel que

$$c_0 n^{1/(2s+1)} \leq 2^{j_3} \leq C_0 n^{1/(2s+1)},$$

où  $c_0$  et  $C_0$  désignent deux constantes choisies a posteriori. Pour toute suite  $\varepsilon = (\varepsilon_k)_{k \in \{0, \dots, 2^{j_3}-1\}} \in \{0, 1\}^{2^{j_3}}$ , on définit  $h_\varepsilon^* : [0, 1] \rightarrow \mathbb{R}$  par

$$h_\varepsilon^*(x) = L_* 2^{-j_3(s+1/2)} \sum_{k=0}^{2^{j_3}-1} \varepsilon_k \psi_{j_3,k}(H(x)), \quad (4.12)$$

#### 4.6. DÉMONSTRATIONS

---

où  $L_* > 0$  est une constante choisie a posteriori. Les coefficients d'ondelettes déformées de  $h_\varepsilon$  valent

$$\beta_{j,k}^H = \int_0^1 h_\varepsilon^*(H^{-1}(x)) \psi_{j,k}(x) dx = \begin{cases} L_* 2^{-j_3(s+1/2)} \varepsilon_k, & \text{si } j = j_3, \\ 0, & \text{sinon,} \end{cases}$$

On a donc  $2^{j_3(s+1/2)} (\sum_{k=0}^{2^{j_3}-1} |\beta_{j_3,k}^H|^\pi w(I_{j_3,k}))^{1/\pi} = L_*$ . Comme  $w \in \mathcal{A}_p \subseteq \mathcal{A}_\pi$ , par un choix convenable de  $L_*$ , on a  $h_\varepsilon^* \in B_{\pi,r}^{s,H}(L)$ .

Le Théorème de Varshamov-Gilbert assure l'existence d'un sous-ensemble  $E_j = \{\varepsilon^{(0)}, \dots, \varepsilon^{(T_{j_3})}\}$  de  $\{0, 1\}^{2^{j_3}}$  et de deux constantes  $c \in ]0, 1[$ ,  $\alpha \in ]0, 1[$  telles que, pour tout  $u \neq v$ ,  $u, v \in \{0, \dots, T_{j_3}\}$ , on a

$$\sum_{k=0}^{2^{j_3}-1} |\varepsilon_k^{(u)} - \varepsilon_k^{(v)}| \geq c 2^{j_3} \quad \text{et} \quad T_{j_3} = \exp(\alpha 2^{j_3}).$$

Grâce à l'inégalité (4.5), pour tout  $u \neq v$  et  $u, v \in \{0, \dots, T_{j_3}\}$ , il vient

$$\begin{aligned} \|h_{\varepsilon^{(u)}}^* - h_{\varepsilon^{(v)}}^*\|_p &\geq c 2^{j_3/2} 2^{-j_3(s+1/2)} \left( \sum_{k=0}^{2^{j_3}-1} |\varepsilon_k^{(u)} - \varepsilon_k^{(v)}|^p w(I_{j_3,k}) \right)^{1/p} \\ &= c 2^{-j_3 s} \left( \sum_{k=0}^{2^{j_3}-1} |\varepsilon_k^{(u)} - \varepsilon_k^{(v)}| w(I_{j_3,k}) \right)^{1/p}. \end{aligned}$$

Il reste à prouver l'existence d'une constante  $c > 0$  telle que

$$\sum_{k=0}^{2^{j_3}-1} |\varepsilon_k^{(u)} - \varepsilon_k^{(v)}| w(I_{j_3,k}) \geq c.$$

Si  $v \asymp 1$  alors c'est une conséquence immédiate du Théorème de Varshamov-Gilbert. Si  $v \not\asymp 1$  alors c'est une conséquence de la propriété du doubling vérifiée par  $w$ . En effet, pour tout  $u \neq v$ ,  $u, v \in E_{j_3}$ , posons  $\mathcal{N}_{u,v} = \{k \in \{0, \dots, 2^{j_3}-1\}, \varepsilon_k^{(u)} \neq \varepsilon_k^{(v)}\}$  et observons que

$$\sum_{k=0}^{2^{j_3}-1} |\varepsilon_k^{(u)} - \varepsilon_k^{(v)}| w(I_{j_3,k}) = \sum_{k \in \mathcal{N}_{u,v}} w(I_{j_3,k}).$$

Par définition de  $E_{j_3}$ , il existe une constante  $c \in ]0, 1[$  telle que

$$\text{Card}(\mathcal{N}_{u,v}) \geq c 2^{j_3}.$$

Trois possibilités s'offrent à nous.



1. Il existe une constante  $c^* \in ]0, 1[$  et deux entiers  $a$  et  $b$  tels que

$$\begin{cases} \sum_{k \in \mathcal{N}_{u,v}} w(I_{j_3,k}) \geq w([0, a2^{-j_3}]) + w([(b+1)2^{-j_3}, 1]), \\ |b-a| = c^* 2^{j_3} - 1. \end{cases}$$

Par la propriété du doubling (4.3) et une inégalité élémentaire de convexité, il existe une constante  $C > 0$  telle que

$$\begin{aligned} \sum_{k \in \mathcal{N}_{u,v}} w(I_{j_3,k}) &\geq w([0, a2^{-j_3}]) + w([(b+1)2^{-j_3}, 1]) \\ &\geq Cw([0, 1])(|a2^{-j_3}|^p + |1 - (b+1)2^{-j_3}|^p) \\ &\geq C|1 - (b-a)2^{-j_3} - 2^{-j_3}|^p = C|1 - c^*|^p. \end{aligned}$$

2. Il existe une constante  $c^* \in ]0, 1[$  et deux entiers  $a$  et  $b$  tels que

$$\begin{cases} \sum_{k \in \mathcal{N}_{u,v}} w(I_{j_3,k}) \geq w([a2^{-j_3}, (b+1)2^{-j_3}]), \\ |b-a| = c^* 2^{j_3} - 1. \end{cases}$$

Par la propriété du doubling (4.3), il existe d'une constante  $C > 0$  telle que

$$\begin{aligned} \sum_{k \in \mathcal{N}_{u,v}} w(I_{j_3,k}) &\geq w([a2^{-j_3}, (b+1)2^{-j_3}]) \\ &\geq C|2^{-j_3} + (b-a)2^{-j_3}|^p = C(c^*)^p. \end{aligned}$$

3. Les espaces entre les  $w(I_{j,k})$  avec  $k \in \mathcal{N}_{u,v}$  sont petits. Dans ce cas, l'hypothèse du doubling nous permet de combler ces espaces. En effet, il existe une constante  $C > 0$  telle que, pour tout entier  $l \in \mathbb{N}^*$  indépendant de  $j$ , on a

$$w(I_{j,k}) = \int_{k2^{-j}}^{(k+1)2^{-j}} w(x)dx \geq c \int_{k2^{-j}}^{(k+l)2^{-j}} w(x)dx = c \sum_{i=0}^{l-1} w(I_{j,k+i}).$$

Ainsi, pour tout  $u \neq v$  et  $u, v \in \{0, \dots, T_{j_3}\}$ , on a

$$\|h_{\varepsilon(u)}^* - h_{\varepsilon(v)}^*\|_p \geq c 2^{j_3/2} 2^{-j_3(s+1/2)} = 2^{-j_3 s} = 2\nu.$$

Par conséquent, les fonctions  $h_{\varepsilon(\cdot)}^*$  vérifient l'hypothèse (C1) avec  $m = \exp(\alpha 2^{j_3})$  et  $\nu \asymp 2^{-j_3 s}$ .

– *Hypothèse (C2).* Soit  $u \neq v$ ,  $u, v \in E_{j_3}$ . Pour tout  $t \in [0, 1]$ , la propriété de concentration (2.5) entraîne

$$|h_{\varepsilon(u)}^*(t) - h_{\varepsilon(v)}^*(t)| \leq C 2^{-j_3(s+1/2)} \sum_{k=0}^{2^{j_3}-1} |\psi_{j_3,k}(H(t))| \leq C 2^{-j_3(s+1/2)} 2^{j_3/2} = C 2^{-j_3 s}.$$

## 4.6. DÉMONSTRATIONS

---

En utilisant la relation (3.8), on a

$$\begin{aligned} \mathbb{E}_{h_{\varepsilon(u)}^*}^n (\log(\Lambda(\mathbb{P}_{h_{\varepsilon(u)}^*}^n, \mathbb{P}_{h_{\varepsilon(v)}^*}^n))) &= 2^{-1}n \int_0^1 |h_{\varepsilon(u)}^*(t) - h_{\varepsilon(v)}^*(t)|^2 v(t)^{-2} dt \\ &\leq C 2^{j_3} n 2^{-j_3(2s+1)} \int_0^1 v(t)^{-2} dt \leq C c_0^{-(2s+1)} 2^{j_3}. \end{aligned}$$

L'hypothèse (C2) est vérifiée en prenant  $c_0$  suffisamment grand.

En appliquant le Théorème 3.2.3 avec

$$v \asymp 2^{-j_3 s} \asymp n^{-\alpha_1}, \quad \alpha_1 = s/(2s+1),$$

on prouve l'existence d'une constante  $c > 0$  telle que

$$\mathcal{R}_{n,p}(B_{\pi,r}^{s,H}(L)) \geq c n^{-\alpha_1 p}.$$

Cela termine la preuve du Théorème 4.4.1. Une preuve alternative utilisant le Lemme d'Assouad est consultable dans l'article de Chesneau (2005).  $\square$

**Preuve du Théorème 4.4.2.** En tenant compte de la sous-section 4.3.2, il suffit de montrer que les hypothèses analogues aux hypothèses (01) et (02) de la sous-section 2.4.2 sont satisfaites.

Par définition de l'estimateur  $\hat{\beta}_{j,k}^H$  et un changement de variables adapté, on a

$$\hat{\beta}_{j,k}^H - \beta_{j,k}^H = n^{-1/2} \int_0^1 \psi_{j,k}(H(t)) v(t)^{-1} dt \sim \mathcal{N}(0, n^{-1}).$$

En utilisant des inégalités gaussiennes élémentaires, pour tout  $j \in \{j_1, \dots, j_2\}$  et  $k \in \{0, \dots, 2^j - 1\}$ , il vient

$$\mathbb{E}_f^n (|\hat{\beta}_{j,k}^H - \beta_{j,k}^H|^{2p}) \leq C n^{-p},$$

et, pour  $\mu$  suffisamment grand,

$$\mathbb{P}_f^n \left( |\hat{\beta}_{j,k}^H - \beta_{j,k}^H| \geq 2^{-1} \mu \sqrt{(\log n/n)} \right) \leq 2 \exp(-8^{-1} \mu^2 \log n) \leq C n^{-p}.$$

Cela termine la preuve du Théorème 4.4.2.  $\square$

**Preuve du Théorème 4.4.3.** Étudions séparément la preuve de la borne supérieure et la preuve de la borne inférieure.

• *Borne supérieure.* Soit  $\pi \geq p$ . Kerkyacharian et Picard (2005, Proposition 3) ont montré que l'hypothèse (D) est satisfaite avec l'estimateur

$$\hat{\alpha}_{j,k}^G = n^{-1} \sum_{i=1}^n Y_i \phi_{j,k}(G(X_i)).$$

Pour  $\pi \geq p$ , le Théorème 4.3.1 entraîne l'existence d'une constante  $C > 0$  telle que

$$\mathcal{R}_{n,p}(B_{\pi,r}^{s,G}(L)) \leq C n^{-\alpha_1 p}, \quad \alpha_1 = s/(2s+1).$$

• *Borne inférieure.* Rappel : pour toutes fonctions  $h_v, h_l : [0, 1] \rightarrow \mathbb{R}$  telles que  $h_v \neq h_l$  et  $\mathbb{P}_{h_v}^n \ll \mathbb{P}_{h_l}^n$ , on a

$$\mathbb{E}_{h_v}^n(\log(\Lambda(\mathbb{P}_{h_v}^n, \mathbb{P}_{h_l}^n))) = 2^{-1} n \int_0^1 |h_v(t) - h_l(t)|^2 g(t) dt. \quad (4.13)$$

En procédant de manière identique à la preuve de la borne inférieure du Théorème 4.4.1 avec  $g$  à la place de  $1/v^2$ , on montre l'existence d'une constante  $c > 0$  telle que

$$\mathcal{R}_{n,p}(B_{\pi,r}^{s,H}(L)) \geq c n^{-\alpha_1 p}, \quad \alpha_1 = s/(2s+1).$$

Cela termine la preuve du Théorème 4.4.3. □

\* \* \*

Chapitre

5

## Seuillage par blocs

Dans ce chapitre, nous évaluons les performances de plusieurs estimateurs en ondelettes reposant sur le seuillage par blocs. Nous considérons les approches minimax et maxiset sous le risque  $\mathbb{L}^p$ . Tous nos résultats sont applicables à de nombreux modèles statistiques, y compris certains problèmes inverses.

### Sommaire

---

<b>5.1</b>	<b>Introduction</b>	<b>84</b>
<b>5.2</b>	<b>Estimateurs et hypothèses</b>	<b>86</b>
5.2.1	Estimateurs de seuillage par blocs	86
5.2.2	Hypothèses	88
<b>5.3</b>	<b>Encadrement du risque</b>	<b>88</b>
5.3.1	Minoration	88
5.3.2	Majoration	89
<b>5.4</b>	<b>Résultats minimax</b>	<b>90</b>
5.4.1	Bornes supérieures	90
5.4.2	Comparaison minimax	92
<b>5.5</b>	<b>Résultats maxisets</b>	<b>94</b>
5.5.1	Maxisets	94
5.5.2	Comparaison maxiset	95
<b>5.6</b>	<b>Applications</b>	<b>96</b>
5.6.1	Bruit blanc gaussien	96
5.6.2	Régression à pas aléatoires	97
5.6.3	Convolution en bruit blanc gaussien	98
5.6.4	Vers la pratique	100
<b>5.7</b>	<b>Démonstrations</b>	<b>103</b>
5.7.1	Preuves des résultats de la Section 5.3	103
5.7.2	Preuves des résultats de la Section 5.4	111
5.7.3	Preuves des résultats de la Section 5.5	118
5.7.4	Preuves des résultats de la Section 5.6	122

---

## 5.1 Introduction

### MOTIVATIONS

Notre objectif est le suivant : à partir d'un modèle statistique général, on veut construire des estimateurs adaptatifs minimisant autant que possible le risque  $\mathbb{L}^p$ . Pour ce faire, nous considérons le seuillage par blocs en ondelettes et tout particulièrement :

- l'estimateur de seuillage global construit par Kerkycharian et al. (1996),
- une version  $\mathbb{L}^p$  de l'estimateur BlockShrink développé par Cai (1996).

L'intérêt que nous portons à ces constructions est détaillé ci-après.

Les notions d'optimalité, de vitesse minimax et de supériorité maxiset utilisées sont sous-entendues "pour de nombreux modèles statistiques".

- *Point de vue minimax.* Voici les motivations et les trois grandes questions auxquelles nous répondons :

L'estimateur de seuillage global est optimal sous le risque  $\mathbb{L}^p$  sur  $B_{\pi,r}^s(L)$  avec  $\pi \geq p$  pour de nombreux modèles statistiques. Voir, par exemple, les articles de Kerkycharian et al. (1996) et Pensky et Vidakovic (1998). Cependant, qu'en est-il si  $p > \pi$  ?

L'estimateur BlockShrink est optimal sous le risque  $\mathbb{L}^2$  sur  $B_{\pi,r}^s(L)$  avec  $\pi \geq 2$  et presque optimal avec  $1 < \pi < 2$  pour de nombreux modèles statistiques. Voir, par exemple, les articles de Cai (1996; 1997; 1999; 2002b) et Chicken (2003b). Cependant, peut-on étendre ce résultat sous le risque  $\mathbb{L}^p$  avec  $p \neq 2$  ?

L'estimateur de seuillage dur est sous-optimal sous le risque  $\mathbb{L}^2$  sur  $B_{\pi,r}^s(L)$  avec  $\pi \geq 2$  pour le modèle de régression à pas équidistants (1.2). Cela a été montré par Cai (2002b, Theorem 1). Cependant, qu'en est-il si on se place sous le risque  $\mathbb{L}^p$  avec  $p \neq 2$  et si on traite de modèles statistiques plus complexes ?

- *Point de vue maxiset.* Les estimateurs de seuillage par blocs ont été étudiés par Cohen et al. (2000b) et Autin (2005).

Entre autre, Autin (2005) a montré que la version  $\mathbb{L}^p$  de l'estimateur BlockShrink est meilleure au sens maxiset que l'estimateur de seuillage dur. Cela est obtenu sous le risque de Besov  $B_{p,p}^0$  avec  $p \geq 2$  pour le modèle de bruit blanc gaussien et les vitesses  $w_n \in \{n^{-\alpha p/2}, (\log n)^{\alpha p/2} n^{-\alpha p/2}\}$  avec  $\alpha \in ]0, 1[$ . Cependant :

- est-ce que l'estimateur BlockShrink considéré est strictement meilleur au sens maxiset que l'estimateur de seuillage dur ?
- est-ce que cela peut-être étendu sous le risque  $\mathbb{L}^p$  pour un large panel de modèles statistiques ?

Nous apportons des réponses précises à ces questions. Précisons que le risque  $\mathbb{L}^p$  est plus naturel à utiliser que le risque  $B_{p,p}^0$ . De plus, si  $p \geq 2$ , pour tout estimateur  $\hat{f}_n$  de  $f$ , on a

$$\mathbb{E}_f^n(\|\hat{f}_n - f\|_{B_{p,p}^0}^p) \leq \mathbb{E}_f^n(\|\hat{f}_n - f\|_p^p).$$

En ce sens, le risque  $\mathbb{L}^p$  est plus puissant que le risque  $B_{p,p}^0$ . En contrepartie, son utilisation fait appel à des outils mathématiques pointus.

Un des problèmes moteurs de nos applications concerne le modèle de régression à pas aléatoires (1.4). Quand la loi des pas est uniforme, Chicken (2003b) a montré l'optimalité de l'estimateur BlockShrink en utilisant l'approche minimax sous le risque  $\mathbb{L}^2$  sur des boules de Hölder. Notre dernier objectif est d'étendre ce résultat sous le risque  $\mathbb{L}^p$  sur des boules de Besov quand la densité des pas est connue et bornée.

## PRÉSENTATION

En premier lieu, nous encadrons le risque  $\mathbb{L}^p$  d'une large famille d'estimateurs de seuillage par blocs sans aucune hypothèse de régularité sur la fonction inconnue. Elles sont obtenues en combinant certaines propriétés géométriques des bases d'ondelettes en norme  $\mathbb{L}^p$  avec un découpage adéquat. Il en découle plusieurs résultats minimax et maxisets. Nous les décrivons ci-après.

D'abord, nous déterminons les vitesses de convergence atteintes par ces estimateurs sur  $B_{\pi,r}^s(L)$  avec  $\pi \geq p$ . Celles-ci sont minimax pour de nombreux modèles statistiques. Nous complétons ce résultat en considérant la classe  $B_{\pi,r}^s(L)$  avec  $p > \pi$  et deux estimateurs de seuillage par blocs bien particuliers. Il s'agit de l'estimateur de seuillage global et une version  $\mathbb{L}^p$  de l'estimateur BlockShrink. Nous montrons la sous-optimalité du premier et l'optimalité quasi-généralisée du second. Une étude annexe concernant la sous-optimalité de l'estimateur de seuillage dur sous le risque  $\mathbb{L}^p$  sur  $B_{\pi,r}^s(L)$  avec  $\pi \geq p$  est faite. Grâce à elle, nous concluons à la supériorité minimax de l'estimateur BlockShrink considéré sur l'estimateur de seuillage dur.

Ensuite, nous isolons le maxiset associé à l'estimateur de seuillage global pour la vitesse de convergence  $w_n = n^{-\alpha p/2}$  avec  $\alpha \in ]0, 1[$ . Ce résultat étend celui de Cohen et al. (2000b, Theorem 5) obtenu sous le risque  $\mathbb{L}^2$ . Puis, nous prouvons que l'estimateur BlockShrink considéré est strictement meilleur au sens maxiset que l'estimateur de seuillage dur pour les vitesses  $w_n \in \{n^{-\alpha p/2}, (\log n)^{\alpha p/2} n^{-\alpha p/2}\}$  avec  $\alpha \in ]0, 1[$ .

Finalement, nous mettons en relief la souplesse de nos résultats en les appliquant au modèle de régression à pas aléatoires (1.5) et au modèle de convolution en bruit blanc gaussien. Ce dernier appartient à la famille des problèmes de déconvolution.

## 5.2 Estimateurs et hypothèses

Dans les sections 5.2 à 5.6, on travaille avec une base d'ondelettes sur l'intervalle  $[0, 1]$ . On suppose qu'elle vérifie la propriété d'inconditionnalité (2.3), la propriété de Temlyakov (2.4) et la propriété élémentaire (2.6). On considère un modèle général à partir duquel on peut estimer les coefficients d'ondelettes  $\alpha_{j,k}$  et  $\beta_{j,k}$  associés à la fonction inconnue  $f : [0, 1] \rightarrow \mathbb{R}$ . Les estimateurs correspondants sont notés  $\hat{\alpha}_{j,k}$  et  $\hat{\beta}_{j,k}$ .

Avant de définir les principaux estimateurs de notre étude, expliquons le rôle des facteurs  $\delta$  et  $\nu$  apparaissant dans nos définitions, hypothèses et résultats. Le premier représente une donnée provenant du modèle statistique abordé. Il joue un rôle crucial dans l'étude de certains problèmes inverses. La plupart du temps, il est réduit à zéro. Le facteur  $\nu$  a une utilité purement technique. Dans notre étude, ces deux paramètres sont choisis indépendamment de la régularité supposée de  $f$ . Tous les estimateurs utilisés dans ce chapitre son adaptatifs.

### 5.2.1 Estimateurs de seuillage par blocs

L'idée du seuillage par blocs a été introduite par Efroimovich (1985) dans le cadre de l'analyse de Fourier. Elle a été adaptée au contexte de l'analyse en ondelettes par Kerkycharian et al. (1996). Les premiers estimateurs de seuillage par blocs localisés ont été développés par Hall et al. (1998; 1999) et Cai (1996; 1997).

#### 5.2.1.1 Construction générale

La construction générale des estimateurs de seuillage par blocs est décrite ci-dessous.

**DÉFINITION (ESTIMATEUR DE SEUILLAGE PAR BLOCS)**

*Soient  $p \in [1, \infty[$ ,  $\mu \in ]0, \infty[$ ,  $\delta \in [0, \infty[$  et  $\nu \in ]0, (2\delta + 1)^{-1}]$ . Soient  $j_1$  et  $j_2$  des entiers tels que*

$$2^{j_1} \asymp (\log n)^{(p/2) \vee 1} \quad \text{et} \quad 2^{j_2} \asymp n^\nu \quad (\text{ou} \quad 2^{j_2} \asymp (n/\log n)^\nu).$$

*Pour tout  $j \in \{j_1, \dots, j_2\}$ , on considère les ensembles*

$$\mathcal{B}_{j,K} = \{k \in \{0, \dots, 2^j - 1\}; \quad (K-1)l_j \leq k \leq Kl_j - 1\}, \quad K \in \mathcal{A}_j,$$

*où  $l_j$  est une suite croissante en  $j$  telle que  $l_{j_1} \asymp (\log n)^{(p/2) \vee 1}$  et*

$$\mathcal{A}_j = \{1, \dots, 2^j l_j^{-1}\}.$$

On a supposé que  $2^j l_j^{-1} \in \mathbb{N}^*$  pour  $j \in \{j_1, \dots, j_2\}$ . On définit l'estimateur de seuillage par blocs  $\hat{f}_n^* : [0, 1] \rightarrow \mathbb{R}$  par

$$\hat{f}_n^*(x) = \sum_{k=0}^{2^{j_1}-1} \hat{\alpha}_{j_1,k} \phi_{j_1,k}(x) + \sum_{j=j_1}^{j_2} \sum_{K \in \mathcal{A}_j} \sum_{k \in \mathcal{B}_{j,K}} \hat{\beta}_{j,k} 1_{\{\hat{b}_{j,K} \geq \mu 2^{\delta j} n^{-1/2}\}} \psi_{j,k}(x), \quad (5.1)$$

avec

$$\hat{b}_{j,K} = (l_j^{-1} \sum_{k \in \mathcal{B}_{j,K}} |\hat{\beta}_{j,k}|^p)^{1/p}.$$

À partir de celle-ci, on distingue l'estimateur de seuillage global introduit par Kerkycharian et al. (1996) et une version  $\mathbb{L}^p$  de l'estimateur BlockShrink développé par Cai (1996). Le second étant une version localisée du premier.

**DÉFINITION (ESTIMATEUR DE SEUILLAGE GLOBAL)**

On définit l'estimateur de seuillage global  $\hat{f}_n^g : [0, 1] \rightarrow \mathbb{R}$  par l'estimateur  $\hat{f}_n^*$  décrit au (5.1) avec

$$l_j = 2^j.$$

**DÉFINITION (ESTIMATEUR BLOCKSHRINK)**

On définit l'estimateur BlockShrink sous-entendu "version  $\mathbb{L}^p$  de la construction proposée par Cai (1996)"  $\hat{f}_n^b : [0, 1] \rightarrow \mathbb{R}$  par l'estimateur  $\hat{f}_n^*$  décrit au (5.1) avec

$$l_j \asymp (\log n)^{(p/2) \vee 1}.$$

La première version  $\mathbb{L}^p$  de l'estimateur BlockShrink est apparue dans l'article de Picard et Tribouley (2000) dans le contexte des intervalles de confiance.

### 5.2.1.2 Estimateur de seuillage dur

Rappelons la définition de l'estimateur de seuillage dur. Notons la présence du facteur  $\delta$  dans l'expression du seuil.

**DÉFINITION (ESTIMATEUR DE SEUILLAGE DUR)**

Soient  $\delta \in [0, \infty[$ ,  $\nu \in ]0, (2\delta + 1)^{-1}]$  et  $\mu \in ]0, \infty[$ . Soient  $j_1$  et  $j_2$  des entiers tels que

$$j_1 = \tau \quad \text{et} \quad 2^{j_2} \asymp (n/\log n)^\nu.$$

On définit l'estimateur de seuillage dur sous-entendu "à seuil universel"  $\hat{f}_n^h : [0, 1] \rightarrow \mathbb{R}$  par

$$\hat{f}_n^h(x) = \sum_{k=0}^{2^{j_1}-1} \hat{\alpha}_{j_1,k} \phi_{j_1,k}(x) + \sum_{j=j_1}^{j_2} \sum_{k=0}^{2^j-1} \hat{\beta}_{j,k} 1_{\{|\hat{\beta}_{j,k}| \geq \mu 2^{\delta j} \sqrt{(\log n/n)}\}} \psi_{j,k}(x). \quad (5.2)$$

Dans la suite, par souci de lisibilité, on adopte les notations

$$\sum_K = \sum_{K \in \mathcal{A}_j} \quad \text{et} \quad \sum_{(K)} = \sum_{k \in \mathcal{B}_{j,K}}.$$

Sauf précision explicite, les notions d'optimalité, de vitesse minimax et de supériorité maxiset seront sous-entendues "pour de nombreux modèles statistiques".



### 5.2.2 Hypothèses

Les hypothèses clés de l'étude sont décrites ci-dessous.

**HYPOTHÈSE (F1)**

Posons  $\hat{\beta}_{j_1-1,k} = \hat{\alpha}_{j_1,k}$ . Il existe une constante  $C > 0$  telle que, pour tout  $j \in \{j_1 - 1, \dots, j_2\}$ ,  $k \in \{0, \dots, 2^j - 1\}$  et  $n$  suffisamment grand, on a

$$\mathbb{E}_f^n(|\hat{\beta}_{j,k} - \beta_{j,k}|^{2p}) \leq C 2^{2\delta j p} n^{-p}.$$

**HYPOTHÈSE (F2)**

Il existe deux constantes  $\mu_1 > 0$  et  $C > 0$  telles que, pour tout  $j \in \{j_1, \dots, j_2\}$ ,  $K \in \mathcal{A}_j$  et  $n$  suffisamment grand, on a

$$\mathbb{P}_f^n((l_j^{-1} \sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^p)^{1/p} \geq 2^{-1} \mu_1 2^{\delta j} n^{-1/2}) \leq C n^{-p}.$$

**HYPOTHÈSE (F3)**

Il existe deux constantes  $\mu_2 > 0$  et  $C > 0$  telles que, pour  $j \in \{j_1, \dots, j_2\}$ ,  $k \in \{0, \dots, 2^j - 1\}$  et  $n$  suffisamment grand, on a

$$\mathbb{P}_f^n(|\hat{\beta}_{j,k} - \beta_{j,k}| \geq 2^{-1} \mu_2 2^{\delta j} \sqrt{(\log n/n)}) \leq C n^{-p}.$$

Le facteur  $p \in [1, \infty[$  correspond à celui du risque  $\mathbb{L}^p$  considéré. Les hypothèses (F1), (F2) et (F3) sont vérifiables pour de nombreux modèles statistiques. Plusieurs applications sont traitées dans la Section 5.6.

## 5.3 Encadrement du risque

Dans cette section, nous encadrons le risque  $\mathbb{L}^p$  associé à l'estimateur de seuillage par blocs  $\hat{f}_n^*$  défini par (5.1) sans aucune hypothèse de régularité sur  $f$ .

### 5.3.1 Minoration

Le Théorème 5.3.1 ci-dessous fournit une minoration du risque  $\mathbb{L}^p$  de l'estimateur de seuillage par blocs  $\hat{f}_n^*$  défini par (5.1).

**THÉORÈME 5.3.1**

Soit  $p \in ]1, \infty[$ . Plaçons nous dans le cadre statistique général décrit au début de la Section 5.2. Supposons que les hypothèses (F1) et (F2) soient satisfaites. Considérons l'estimateur de seuillage par blocs  $\hat{f}_n^*$  défini par (5.1) avec la constante de seuillage  $\mu = \mu_1$ . Alors il existe une constante  $C > 0$  telle que, pour tout  $\alpha \in ]0, 1[$  et  $n$  suffisamment grand, on a

$$\mathcal{Q}_1(f) + \mathcal{Q}_2(f) \leq C(\mathbb{E}_f^n(\|\hat{f}_n^* - f\|_p^p) + n^{-\alpha p/2}),$$

où

$$\mathcal{Q}_1(f) = \left\| \sum_{j=\tau}^{\infty} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{b_{j,K} \leq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} \psi_{j,k} \right\|_p^p$$

et

$$\mathcal{Q}_2(f) = \left\| \sum_{j=j_2+1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k} \right\|_p^p.$$

La preuve utilise des techniques de calculs similaires à celles développées dans l'article de Kerkycharian et Picard (2000, Theorem 5.1). Les propriétés géométriques de la base d'ondelettes y jouent un rôle déterminant.

L'intérêt du Théorème 5.3.1 est de fournir une minoration du risque dépendant de la nature de  $\hat{f}_n^*$ . Cela est utile pour étudier la sous-optimalité de certains estimateurs de seuillage par blocs ainsi que le maxiset associé.

REMARQUE : Dans le cadre du modèle de bruit blanc gaussien, des résultats proches du Théorème 5.3.1 ont été établis sous le risque  $\mathbb{L}^2$ . Voir, par exemple, les articles de Cai (2002b, Theorem 1) et Autin (2005, Theorem 3.1).

### 5.3.2 Majoration

Rappelons que la norme de Besov  $B_{p,p}^0$  est définie par

$$\|f\|_{B_{p,p}^0}^p \asymp \sum_{k=0}^{2^\tau-1} |\alpha_{\tau,k}|^p + \sum_{j=\tau}^{\infty} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p.$$

Le Théorème 5.3.2 ci-dessous fournit une majoration du risque  $\mathbb{L}^p$  de l'estimateur de seuillage par blocs  $\hat{f}_n^*$  défini par (5.1).

#### THÉORÈME 5.3.2

Soit  $p \in [1, \infty[$ . Plaçons nous dans le cadre statistique général décrit au début de la Section 5.2. Supposons que les hypothèses (F1) et (F2) soient satisfaites. Considérons l'estimateur de seuillage par blocs  $\hat{f}_n^*$  défini par (5.1) avec la constante de seuillage  $\mu = \mu_1$ . Alors il existe une constante  $C > 0$  telle que, pour tout  $\alpha \in ]0, 1[$  et  $n$  suffisamment grand, on a

$$\mathbb{E}_f^n(\|\hat{f}_n^* - f\|_p^p) \leq C(\mathcal{Q}_1^*(f) + \mathcal{Q}_2(f) + n^{-\alpha p/2}),$$

où

$$\mathcal{Q}_1^*(f) = \sum_{m \in \mathbb{N}} 2^{-mp} \left\| \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{b_{j,K} \leq 2^{-1} \mu_1 n^{-1/2} 2^{\delta j} 2^{m+1}\}} \psi_{j,k} \right\|_u^p,$$

avec  $u = B_{p,p}^0$  si  $p \in ]1, 2]$  et  $u = p$  si  $p \in [2, \infty[$ . Le terme  $\mathcal{Q}_2(f)$  est défini comme dans le Théorème 5.3.1.

La preuve repose sur des outils mathématiques similaires à ceux utilisés dans l'article de Kerkycharian et Picard (2000, Theorem 5.2). Là encore, les propriétés géométriques de la base d'ondelettes vont être déterminantes.

Les nouveautés du Théorème 5.3.1 et 5.3.2 résident dans la considération du risque  $\mathbb{L}^p$  avec  $p > 1$  et dans la souplesse concernant le modèle statistique initial. Ils sont le fer de lance de nos résultats minimax et maxisets à venir.

## 5.4 Résultats minimax

L'objectif de cette section est

- d'évaluer les performances de plusieurs estimateurs de seuillage par blocs sous le risque  $\mathbb{L}^p$  sur  $B_{\pi,r}^s(L)$ ,
- de faire une étude comparative entre ceux-ci et l'estimateur de seuillage dur.

### 5.4.1 Bornes supérieures

Cette sous-section est divisée en deux parties. La première partie est consacrée aux bornes supérieures de l'estimateur de seuillage par blocs sous le risque  $\mathbb{L}^p$  sur  $B_{\pi,r}^s(L)$  avec  $\pi \geq p$ . La seconde partie est consacrée aux bornes supérieures de l'estimateur de seuillage global et l'estimateur BlockShrink sous le risque  $\mathbb{L}^p$  sur  $B_{\pi,r}^s(L)$  avec  $p \geq \pi$ .

#### 5.4.1.1 Cas où $\pi \geq p$

Le Théorème 5.4.1 ci-dessous détermine la borne supérieure de l'estimateur de seuillage par blocs  $\hat{f}_n^*$  défini par (5.1) sous le risque  $\mathbb{L}^p$  sur  $B_{\pi,r}^s(L)$  avec  $\pi \geq p$ .

##### THÉORÈME 5.4.1

*Soit  $p \in [1, \infty[$ . Plaçons nous dans le cadre statistique général décrit au début de la Section 5.2. Supposons que les hypothèses (F1) et (F2) soient satisfaites. Considérons l'estimateur de seuillage par blocs  $\hat{f}_n^*$  défini par (5.1) avec la constante de seuillage  $\mu = \mu_1$ . Alors il existe une constante  $C > 0$  telle que, pour tout  $\pi \in [p, \infty]$ ,  $s \in ]1/\pi - (1/2 - 1/(2\nu) + \delta)_+, N]$ ,  $r \in [1, \infty]$  et  $n$  suffisamment grand, on a*

$$\sup_{f \in B_{\pi,r}^s(L)} \mathbb{E}_f^n(\|\hat{f}_n^* - f\|_p^p) \leq Cn^{-\alpha_1 p},$$

où  $\alpha_1 = s/(2(s + \delta) + 1)$ .

La preuve est la conséquence du Théorème 5.3.2.

La vitesse de convergence obtenue est optimale pour de nombreux modèles statistiques.

REMARQUE : Bien entendu, le Théorème 5.4.1 est applicable avec  $\hat{f}_n^* \in \{\hat{f}_n^g, \hat{f}_n^b\}$ . Il englobe ainsi une grande partie des résultats traitant de l'optimalité minimax de l'estimateur BlockShrink et de l'estimateur de seuillage global sous le risque  $\mathbb{L}^2$  sur  $B_{\pi,r}^s(L)$  avec  $\pi \geq 2$ . Voir, par exemple, les articles de Cai (1996; 1997; 1999; 2002b) pour le modèle de régression à pas équidistants et le modèle de bruit blanc gaussien, de Chicken (2003a) pour le modèle de régression à pas non équidistants, de Kerkycharian et al. (1996), Hall et al. (1999), Pensky (1999) et Cai et Chicken (2005) pour le modèle de densités, de Li et Xiao (2004) pour le modèle de régression à mémoire longue, de Pensky et Vidakovic (1998) pour le modèle de convolution de densités et de Cai (2002a) pour certains problèmes linéaires inverses.

Pour compléter ce résultat, les deux paragraphes suivants étudient les performances minimax l'estimateur de seuillage global et l'estimateur BlockShrink sur  $B_{\pi,r}^s(L)$  avec  $p > \pi$ .

#### 5.4.1.2 Cas où $p \geq \pi$

##### Estimateur de seuillage global

La Proposition 5.4.1 ci-dessous isole une borne inférieure de l'estimateur de seuillage global sous le risque  $\mathbb{L}^p$  sur  $B_{\pi,r}^s(L)$  avec  $p > \pi$ .

##### PROPOSITION 5.4.1

*Soit  $p \in ]1, \infty[$ . Plaçons nous dans le cadre statistique général décrit au début de la Section 5.2. Supposons que les hypothèses (F1) et (F2) soient satisfaites. Considérons l'estimateur de seuillage global  $\hat{f}_n^g$  défini par (5.1) avec  $l_j = 2^j$  et la constante de seuillage  $\mu = \mu_1$ . Alors il existe une constante  $c > 0$  telle que, pour tout  $\pi \in [1, p]$ ,  $s \in ]1/\pi - (1/2 - 1/(2\nu) + \delta)_+, N]$ ,  $r \in [1, \infty]$  et  $n$  suffisamment grand, on a*

$$\sup_{f \in B_{\pi,r}^s(L)} \mathbb{E}_f^n(\|\hat{f}_n^g - f\|_p^p) \geq cn^{-\alpha_3 p},$$

où  $\alpha_3 = (s - 1/\pi + 1/p)/(2(s - 1/\pi + 1/p + \delta) + 1)$ .

La preuve de la Proposition 5.4.1 repose sur le Théorème 5.3.1.

La vitesse de convergence déterminée dans la Proposition 5.4.1 est plus mauvaise que les vitesses minimax généralement obtenues sur  $B_{\pi,r}^s(L)$  avec  $p > \pi$ . En effet, pour  $n$  grand, on a  $((\log n)^{1_{\{\varepsilon \leq 0\}}} / n)^{(\alpha_1 \wedge \alpha_2)p} < Cn^{-\alpha_3 p}$  où  $\alpha_1 = s/(2(s + \delta) + 1)$  et  $\alpha_2 = (s - 1/\pi + 1/p)/(2(s - 1/\pi + \delta) + 1)$ .

REMARQUE : À notre connaissance, ce résultat n'a pas d'antécédent. Dans le même esprit, Cai (2002b, Theorem 1) a montré que si  $0 \leq s < 1$  alors l'estimateur de seuillage par blocs défini avec  $l_j \asymp (\log n)^s$  est sous-optimal sous le risque  $\mathbb{L}^2$  sur des boules de Hölder pour le modèle régression à pas équidistants.

### Estimateur BlockShrink

Le Théorème 5.4.2 ci-dessous s'intéresse aux bornes supérieures de l'estimateur BlockShrink sous le risque  $\mathbb{L}^p$  sur  $B_{\pi,r}^s(L)$  avec  $p > \pi$ .

#### THÉORÈME 5.4.2

Soit  $p \in ]1, \infty[$ . Plaçons nous dans le cadre statistique général décrit au début de la Section 5.2. Supposons que les hypothèses (F1) et (F2) soient satisfaites. Considérons l'estimateur BlockShrink  $\hat{f}_n^b$  défini par (5.1) avec  $l_j \asymp (\log n)^{(p/2)\vee 1}$  et la constante de seuillage  $\mu = \mu_1$ . Alors il existe une constante  $C > 0$  telle que, pour tout  $\pi \in [1, p]$ ,  $s \in ]1/\pi - (1/2 - 1/(2\nu) + \delta)_+, N]$ ,  $r \in [1, \infty]$  et  $n$  suffisamment grand, on a

$$\sup_{f \in B_{\pi,r}^s(L)} \mathbb{E}_f^n(\|\hat{f}_n^b - f\|_p^p) \leq C\varphi_n,$$

où

$$\varphi_n = \begin{cases} (\log n/n)^{\alpha_1 p}, & \text{si } \epsilon > 0, \\ (\log n/n)^{\alpha_2 p} (\log n)^{(p-\pi/r)+1_{\{\epsilon=0\}}}, & \text{si } \epsilon \leq 0, \end{cases}$$

avec  $\alpha_1 = s/(2(s + \delta) + 1)$ ,  $\alpha_2 = (s - 1/\pi + 1/p)/(2(s - 1/\pi + \delta) + 1)$  et  $\epsilon = \pi s + (\delta + 1/2)(\pi - p)$ .

La preuve du Théorème 5.4.2 découle du Théorème 5.3.2.

L'estimateur BlockShrink atteint une vitesse de convergence généralement minimax sur  $B_{\pi,r}^s(L)$  avec  $\epsilon \leq 0$ . La localisation et la taille des blocs sont les clés de cette optimalité.

REMARQUE : La vitesse de convergence obtenue pour le cas  $p > \pi$  et  $\epsilon > 0$  peut être améliorée. En considérant le risque  $\mathbb{L}^2$  et le modèle de régression à pas équidistants, Cai (1997, Theorem 3) a montré que l'estimateur BlockShrink atteint la vitesse  $\varphi_n = n^{-2\alpha_1}(\log n)^{2\alpha_*}$  avec  $\alpha_* = (2/\pi - 1)/(2s + 1)$ .

À notre connaissance, les Théorèmes 5.4.1 et 5.4.2 sont les premiers traitant des performances de l'estimateur BlockShrink sous le risque  $\mathbb{L}^p$  pour  $p \neq 2$  et  $p > 1$ .

### 5.4.2 Comparaison minimax

L'objectif de cette sous-section est de comparer les performances minimax de trois estimateurs adaptatifs en ondelettes. Il s'agit de l'estimateur BlockShrink, de l'estimateur de seuillage global et de l'estimateur de seuillage dur.

### 5.4.2.1 Sous-optimalité de l'estimateur de seuillage dur

La Proposition 5.4.2 ci-dessous prouve que la vitesse de convergence atteinte par l'estimateur de seuillage dur sous les hypothèses (F1) et (F3) sur  $B_{\pi,r}^s(L)$  avec  $\pi \geq p$  ne peut pas être améliorée. Pour plus de précisions, voir le Théorème 2.4.1.

#### PROPOSITION 5.4.2

Soit  $p \in ]1, \infty[$ . Plaçons nous dans le cadre statistique général décrit au début de la Section 5.2. Supposons que les hypothèses (F1) et (F3) soient satisfaites. Considérons l'estimateur de seuillage dur  $\hat{f}_n^h$  défini par (5.2) avec la constante de seuillage  $\mu = \mu_2$ . Alors il existe une constante  $c > 0$  telle que, pour tout  $\pi \in [1, \infty]$ ,  $s \in ]1/\pi - (1/2 - 1/(2\nu) + \delta)_+, N]$ ,  $r \in [1, \infty]$ ,  $\pi s + (\delta + 1/2)(\pi - p) > 0$  et  $n$  suffisamment grand, on a

$$\sup_{f \in B_{\pi,r}^s(L)} \mathbb{E}_f^n(\|\hat{f}_n^h - f\|_p^p) \geq c(\log n/n)^{\alpha_1 p},$$

où  $\alpha_1 = s/(2(s + \delta) + 1)$ .

La preuve repose sur des outils mathématiques similaires à ceux utilisés dans les articles de Picard et Kerkycharian (2000, Theorem 5.2) et Cai (2002b, Theorem 1).

L'intérêt de la Proposition 5.4.2 est de montrer les limites de l'estimateur de seuillage dur dans un contexte statistique très large. Par la même occasion, il justifie notre intérêt pour les estimateurs de seuillage par blocs.

REMARQUE : Lorsque le modèle de régression à pas équidistants est considéré, Cai (2002b, Theorem 1) a montré la Proposition 5.4.2 avec  $p = 2$ ,  $\delta = 0$ ,  $\nu = 1$  et des boules de Hölder.

### 5.4.2.2 Supériorité de l'estimateur BlockShrink

On reprend le cadre statistique posé dans le Théorème 5.4.1, la Proposition 5.4.1, le Théorème 5.4.2 et la Proposition 5.4.2.

- Si  $\pi \geq p$  et  $\hat{f}_n \in \{\hat{f}_n^b, \hat{f}_n^g\}$  alors il existe une constante  $C > 0$  telle que

$$\sup_{f \in B_{\pi,r}^s(L)} \mathbb{E}_f^n(\|\hat{f}_n - f\|_p^p) \leq Cn^{-\alpha_1 p} < C(\log n/n)^{\alpha_1 p} \leq \sup_{f \in B_{\pi,r}^s(L)} \mathbb{E}_f^n(\|\hat{f}_n^h - f\|_p^p).$$

- Si  $p > \pi$  et  $\hat{f}_n \in \{\hat{f}_n^b, \hat{f}_n^h\}$  alors il existe une constante  $C > 0$  telle que

$$\begin{aligned} \sup_{f \in B_{\pi,r}^s(L)} \mathbb{E}_f^n(\|\hat{f}_n - f\|_p^p) &\leq C(\log n/n)^{(\alpha_1 \wedge \alpha_2)p} (\log n)^{(p-\pi/r)+1_{\{\epsilon=0\}}} < Cn^{-\alpha_3 p} \\ &\leq \sup_{f \in B_{\pi,r}^s(L)} \mathbb{E}_f^n(\|\hat{f}_n^g - f\|_p^p). \end{aligned}$$

On observe que l'estimateur BlockShrink est meilleur au sens minimax que l'estimateur de seuillage dur et l'estimateur de seuillage global.

Remarquons que la vitesse atteinte par l'estimateur de seuillage global pour le cas  $p > \pi$  est vraiment plus lente que celle atteinte par l'estimateur de seuillage dur. De manière générale, on peut conclure que l'estimateur de seuillage dur est meilleur au sens minimax que l'estimateur de seuillage global.

TAB. 5.1 – Types d'optimalités généralement constatés pour les trois estimateurs étudiés dans cette section.

Estimateurs	$B_{\pi,r}^s(L)$		
	$\pi \geq p$	$p > \pi$ et $\epsilon > 0$	$\epsilon < 0$
Seuillage dur	Presque optimal*	Presque optimal*	Optimal
Seuillage global	Optimal	Sous-optimal	Sous-optimal
BlockShrink	Optimal	Presque optimal	Optimal

Rappel :  $\epsilon = \pi s + (\delta + 2^{-1})(\pi - p)$ .

\*Cet estimateur est 'exactement' presque optimal.

## 5.5 Résultats maxisets

Cette section est consacrée à l'étude maxiset de l'estimateur BlockShrink et de l'estimateur de seuillage global sous le risque  $\mathbb{L}^p$ .

### 5.5.1 Maxisets

Le Théorème 5.5.1 ci-dessous isole le maxiset associé à l'estimateur de seuillage global sous le risque  $\mathbb{L}^p$  pour la vitesse  $w_n = n^{-\alpha p/2}$  avec  $\alpha \in ]0, 1[$ .

#### THÉORÈME 5.5.1

Soit  $p \in [2, \infty[$ . Plaçons nous dans le cadre statistique général décrit au début de la Section 5.2. Supposons que les hypothèses (F1), (F2) et (F3) soient satisfaites. Considérons l'estimateur de seuillage global  $\hat{f}_n^g$  défini par (5.1) avec  $l_j = 2^j$  et la constante de seuillage  $\mu = \mu_1$ . Alors, pour tout  $\alpha \in ](1 - \nu(2\delta + 1))_+, 1[$ , on a

$$\mathcal{M}(\hat{f}_n^g, p, n^{-\alpha p/2}) = \mathcal{W}_\delta^*((1 - \alpha)p, p).$$

Ici,  $\mathcal{W}_\delta^*((1 - \alpha)p, p)$  désigne l'ensemble des fonctions  $f$  de  $\mathbb{L}^p([0, 1])$  tel que

$$\sup_{u>0} u^{-\alpha p} \left\| \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} 1_{\{b_j \leq u 2^{\delta j}\}} \psi_{j,k} \right\|_p^p < \infty, \quad b_j = (2^{-j} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p)^{1/p}.$$

La preuve du Théorème 5.5.1 repose sur l'encadrement déduit des Théorèmes 5.3.1 et 5.3.2.

REMARQUE : Le Théorème 5.5.1 est la version  $\mathbb{L}^p$  d'un résultat montré par Cohen et al. (2000b, Theorem 5).

### 5.5.2 Comparaison maxiset

Le maxiset associé à l'estimateur BlockShrink est difficile à exprimer à cause de la présence de l'entier  $n$  dans la taille des blocs. Toutefois, grâce aux Théorèmes 5.3.1 et 5.3.2, nous sommes en mesure de le comparer avec d'autres.

Le Théorème 5.5.2 compare le maxiset associé à l'estimateur BlockShrink et l'estimateur de seuillage dur pour la vitesse  $w_n = n^{-\alpha p/2}$  avec  $\alpha \in ]0, 1[$ .

#### THÉORÈME 5.5.2

*Soit  $p \in [2, \infty[$ . Plaçons nous dans le cadre statistique général décrit au début de la Section 5.2. Supposons que les hypothèses (F1), (F2) et (F3) soient satisfaites. Considérons*

- *l'estimateur BlockShrink  $\hat{f}_n^b$  défini par (5.1) avec la constante de seuillage  $\mu = \mu_1$ ,*
- *l'estimateur de seuillage dur  $\hat{f}_n^h$  défini par (5.2) avec la constante de seuillage  $\mu = \mu_2$ .*

*Alors, pour tout  $\alpha \in ](1 - \nu(2\delta + 1))_+, 1[$ , on a*

$$\mathcal{M}(\hat{f}_n^h, p, n^{-\alpha p/2}) \subset \mathcal{M}(\hat{f}_n^b, p, n^{-\alpha p/2}).$$

*Ainsi, l'estimateur BlockShrink considéré est strictement meilleur au sens maxiset que l'estimateur de seuillage dur.*

La preuve repose sur des outils mathématiques similaires à ceux utilisés dans l'article de Kerkycharian et Picard (2000, Theorems 5.1 et 5.2). Pour montrer l'inclusion stricte, nous considérons une fonction de  $\mathcal{M}(\hat{f}_n^b, p, n^{-\alpha p/2})$  qui n'appartient pas à  $\mathcal{M}(\hat{f}_n^h, p, n^{-\alpha p/2})$ .

L'inclusion maxiset du Théorème 5.5.2 est toujours valable si on prend une vitesse de la forme  $w_n = (\log n/n)^{\alpha p/2}$  avec  $\alpha \in ]0, 1[$ .

REMARQUE : Le fait que l'estimateur BlockShrink est meilleur au sens maxiset que l'estimateur de seuillage dur a été montré par Autin (2006, Proposition 4.2) sous le risque  $B_{p,p}^0$  avec  $p \geq 2$  pour le modèle de bruit blanc gaussien. Notre apport personnel réside dans l'utilisation du risque  $\mathbb{L}^p$ , la considération d'un modèle général incluant certains problèmes inverses et la démonstration de la supériorité maxiset stricte.



## 5.6 Applications

Dans cette section, nous illustrons la souplesse de nos résultats en considérant le modèle de bruit blanc gaussien, le modèle de régression à pas aléatoires et le modèle de convolution en bruit blanc gaussien. Nos contributions concernent les deux derniers.

### 5.6.1 Bruit blanc gaussien

On travaille avec la base d'ondelettes  $\zeta$  décrite par (2.1).

#### 5.6.1.1 Résultat principal

Dans le cadre du modèle de bruit blanc gaussien (1.3), les éléments  $\hat{\alpha}_{j,k}$ ,  $\hat{\beta}_{j,k}$ ,  $\delta$ ,  $\nu$ ,  $\mu_1$  et  $\mu_2$  vérifiant les hypothèses (F1), (F2) et (F3) ont été exhibés par Picard et Tribouley (2000).

LEMME 5.6.1 (PICARD ET TRIBOULEY (2000))

*Considérons le modèle de bruit blanc gaussien défini par (1.3). Les hypothèses (F1), (F2) et (F3) sont satisfaites avec*

$$\hat{\alpha}_{j,k} = \int_0^1 \phi_{j,k}(t) dY(t), \quad \hat{\beta}_{j,k} = \int_0^1 \psi_{j,k}(t) dY(t),$$

$\delta = 0$ ,  $\nu = 1$  et  $\mu_1$ ,  $\mu_2$  suffisamment grands.

Si on configure les estimateurs de seuillage par blocs avec les quantités définies dans la Proposition 5.6.1 alors on peut appliquer tous les résultats des sections précédentes.

#### 5.6.1.2 Récapitulatif minimax concernant l'estimateur BlockShrink

Dans le cadre du modèle de bruit blanc gaussien défini par (1.3), le Théorème 5.6.1 synthétise les bornes supérieures de l'estimateur BlockShrink sous le risque  $\mathbb{L}^p$  sur des boules de Besov. C'est une conséquence des Théorèmes 5.4.1 et 5.4.2.

THÉORÈME 5.6.1 (CONSÉQUENCE DES THÉORÈMES 5.4.1 ET 5.4.2)

*Soit  $p \in [1, \infty[$ . Considérons le modèle de bruit blanc gaussien défini par (1.3). Configurons l'estimateur BlockShrink  $\hat{f}_n^b$  défini par (5.1) avec  $l_j \asymp (\log n)^{(p/2) \vee 1}$ , les éléments décrits à la Proposition 5.6.1 et la constante de seuillage  $\mu = \mu_1$ . Alors il existe une constante  $C > 0$  telle que, pour tout  $\pi \in [1, \infty]$   $s \in ]1/\pi, N]$ ,  $r \in [1, \infty]$  et  $n$  suffisamment grand, on a*

$$\sup_{f \in B_{\pi, r}^s(L)} \mathbb{E}_f^n(\|\hat{f}_n^b - f\|_p^p) \leq C\varphi_n,$$

où

$$\varphi_n = \begin{cases} n^{-\alpha_1 p} (\log n)^{\alpha_1 p 1_{\{p > \pi\}}}, & \text{si } \epsilon > 0, \\ (\log n/n)^{\alpha_2 p} (\log n)^{(p-\pi/r)+1_{\{\epsilon=0\}}}, & \text{si } \epsilon \leq 0, \end{cases}$$

## 5.6. APPLICATIONS

---

avec  $\alpha_1 = s/(2s+1)$ ,  $\alpha_2 = (s - 1/\pi + 1/p)/(2(s - 1/\pi) + 1)$  et  $\epsilon = \pi s + 2^{-1}(\pi - p)$ .

L'estimateur  $\hat{f}_n^b$  est optimal, sauf pour le cas  $\epsilon > 0$  avec  $p > \pi$ .

REMARQUE : Le Théorème 5.6.1 est une version  $\mathbb{L}^p$  de deux résultats établis par Cai (1997, Theorems 2 et 3).

### 5.6.2 Régression à pas aléatoires

On travaille de nouveau avec la base d'ondelettes  $\zeta$  décrite par (2.1).

#### 5.6.2.1 Résultat principal

Dans le cadre du modèle de régression à pas aléatoires défini par (3.6), le Théorème 5.6.2 ci-dessous pose les conditions et les éléments  $\hat{\alpha}_{j,k}$ ,  $\hat{\beta}_{j,k}$ ,  $\delta$ ,  $j_2$ ,  $\mu_1$  et  $\mu_2$  vérifiant les hypothèses (F1), (F2) et (F3).

##### THÉORÈME 5.6.2

*Considérons le modèle de régression à pas aléatoires défini par (3.6). Supposons que  $g$  soit connue, bornée inférieurement et bornée supérieurement. Alors les hypothèses (F1), (F2) et (F3) sont vérifiées avec*

$$\hat{\alpha}_{j,k} = n^{-1} \sum_{i=1}^n Y_i g(X_i)^{-1} \phi_{j,k}(X_i), \quad \hat{\beta}_{j,k} = n^{-1} \sum_{i=1}^n Y_i g(X_i)^{-1} \psi_{j,k}(X_i),$$

$\delta = 0$ ,  $j_2$  tel que  $2^{j_2} \asymp \sqrt{n/\log n}$  et  $\mu_1, \mu_2$  suffisamment grands.

La principale difficulté réside dans la démonstration de l'hypothèse (F2). Pour ce faire, on utilise plusieurs techniques de grande déviation telles que l'inégalité de Talagrand (1994) et l'inégalité de Cirelson et al. (1976).

Si on configure les estimateurs de seuillage par blocs avec les quantités définies dans le Théorème 5.6.2 alors on peut appliquer tous les résultats des sections précédentes.

#### 5.6.2.2 Récapitulatif minimax concernant l'estimateur BlockShrink

Dans le cadre du modèle de régression à pas aléatoires défini par (3.6), le Théorème 5.6.3 synthétise les bornes supérieures de l'estimateur BlockShrink sous le risque  $\mathbb{L}^p$  sur des boules de Besov. C'est une conséquence des Théorèmes 5.4.1 et 5.4.2.

##### THÉORÈME 5.6.3 (CONSÉQUENCE DES THÉORÈMES 5.4.1 ET 5.4.2)

*Soit  $p \in [1, \infty[$ . Considérons le modèle de régression à pas aléatoires défini par (3.6). Configurons l'estimateur BlockShrink  $\hat{f}_n^b$  défini par (5.1) avec  $l_j \asymp (\log n)^{(p/2) \vee 1}$ , les éléments décrits au Théorème 5.6.2 et la constante de seuillage  $\mu = \mu_1$ . Alors il existe*

une constante  $C > 0$  telle que, pour tout  $\pi \in [1, \infty]$ ,  $s \in ]1/\pi + 1/2, N]$ ,  $r \in [1, \infty]$  et  $n$  suffisamment grand, on a

$$\sup_{f \in B_{\pi, r}^s(L)} \mathbb{E}_f^n(\|\hat{f}_n^b - f\|_p^p) \leq C\varphi_n,$$

où

$$\varphi_n = \begin{cases} n^{-\alpha_1 p} (\log n)^{\alpha_1 p 1_{\{p > \pi\}}}, & \text{si } \epsilon > 0, \\ (\log n/n)^{\alpha_2 p} (\log n)^{(p-\pi/r)+1_{\{\epsilon=0\}}}, & \text{si } \epsilon \leq 0, \end{cases}$$

avec  $\alpha_1 = s/(2s+1)$ ,  $\alpha_2 = (s - 1/\pi + 1/p)/(2(s - 1/\pi) + 1)$  et  $\epsilon = \pi s + 2^{-1}(\pi - p)$ .

L'estimateur  $\hat{f}_n^b$  est optimal, sauf pour le cas  $\epsilon > 0$  avec  $p > \pi$ .

REMARQUE : Le Théorème 5.4.2 étend un résultat montré par Chicken (2003b, Theorem 2). Ce dernier a été établi sous le risque  $\mathbb{L}^2$  sur des boules de Hölder pour la densité  $g$  uniforme.

### 5.6.3 Convolution en bruit blanc gaussien

#### 5.6.3.1 Description

MODÈLE (CONVOLUTION EN BRUIT BLANC GAUSSIEN)

On observe le processus continu  $\{Y(t); t \in [0, 1]\}$  défini par l'équation stochastique

$$dY(t) = (f \star g)(t)dt + n^{-1/2}dW(t), \quad (5.3)$$

où  $f : [0, 1] \rightarrow \mathbb{R}$  est une fonction inconnue bornée supérieurement,  $g : [0, 1] \rightarrow \mathbb{R}$  est une fonction connue telle que  $g \in \mathbb{L}^1([0, 1])$  et  $\{W(t); t \in [0, 1]\}$  est un mouvement Brownien standard. On suppose que  $f$  et  $g$  sont périodiques sur  $[0, 1]$ . On a posé  $(f \star g)(t) = \int_0^1 f(t-u)g(u)du$ .

L'objectif est de reconstruire  $f$  à partir des observations  $\{\int_0^1 h(t)dY(t), h \in \mathbb{L}^2([0, 1])\}$ .

Le modèle (5.3) a fait l'objet de nombreuses études. Citons Ermakov (1989), Donoho et Low (1992), Koo (1993), Korostelev et Tsybakov (1993), Donoho (1995), Johnstone (1999), Cavalier et Tsybakov (2002) et Johnstone et al. (2004).

#### 5.6.3.2 Ondelettes de Meyer

Dans cette sous-section, on travaille avec une périodisation de la base d'ondelettes de Meyer adaptée à l'intervalle  $[0, 1]$ . Elle est notée

$$\zeta^m = \{\phi_{\tau, k}^m(x), k = 0, \dots, 2^\tau - 1; \psi_{j, k}^m(x); j = \tau, \dots, \infty, k = 0, \dots, 2^j - 1\},$$

où  $\tau$  désigne un entier suffisamment grand. La particularité de celle-ci est que les transformées de Fourier associées aux fonctions  $\psi^m$  et  $\phi^m$  sont à support compact. Pour plus de détails, voir l'ouvrage de Walter (1994).

De plus,  $\zeta^m$  vérifie la propriété d'inconditionnalité pour la norme  $\mathbb{L}^p$  ainsi que la propriété de Temlyakov. Cette dernière a été mise en relief par Johnstone et al. (2004). Pour finir, précisons que les boules de Besov s'expriment en fonction de  $\zeta^m$  de la même manière que pour  $\zeta$ .

### 5.6.3.3 Résultat principal

Dans le cadre du modèle de déconvolution défini par (5.3), le Théorème 5.6.4 ci-dessous pose les conditions et les éléments  $\hat{\alpha}_{j,k}$ ,  $\hat{\beta}_{j,k}$ ,  $\delta$ ,  $\nu$ ,  $\mu_1$  et  $\mu_2$  vérifiant les hypothèses (F1), (F2) et (F3). Ils sont identiques à ceux considérés dans l'article de Johnstone et al. (2004) dans le cadre du seuillage dur.

#### THÉORÈME 5.6.4

*Considérons le modèle de convolution en bruit blanc gaussien défini par (5.3). Supposons que  $g$  vérifie*

$$\mathcal{F}(g)(l) \asymp |l|^{-\delta}, \quad l \in \mathbb{Z}^*, \quad \mathcal{F}(g)(0) \asymp 1. \quad (5.4)$$

*Alors les hypothèses (F1), (F2) et (F3) sont satisfaites avec*

$$\hat{\alpha}_{j,k} = \sum_{l \in C_j} \mathcal{F}^*(Y)(l) \mathcal{F}(g)(l)^{-1} \mathcal{F}(\phi_{j,k}^m)(l), \quad \hat{\beta}_{j,k} = \sum_{l \in C_j} \mathcal{F}^*(Y)(l) \mathcal{F}(g)(l)^{-1} \mathcal{F}(\psi_{j,k}^m)(l),$$

*$\nu = (1 + 2\delta)^{-1}$  et  $\mu_1, \mu_2$  suffisamment grands. On a posé :*

- $C_j = \{l \in \mathbb{Z}; \mathcal{F}(\psi_{j,k}^m)(l) \neq 0\} = \{l \in \mathbb{Z}; |l| \in [2\pi 3^{-1} 2^j, 8\pi 3^{-1} 2^j]\},$
- *pour tout processus  $\{R(t); t \in [0, 1]\}$  vérifiant  $\mathbb{E}(|R(t)|) < \infty$ , on définit l'opérateur  $\mathcal{F}^*(R)$  par  $\mathcal{F}^*(R)(l) = \int_0^1 \exp(-2i\pi lt) dR(t).$*

Là encore, la principale difficulté réside dans la démonstration de l'hypothèse (F2). Pour ce faire, on utilise l'inégalité de Cirelson et al. (1976).

REMARQUE : À titre d'exemple, la fonction  $g : [0, 1] \rightarrow ]0, \infty[$  définie par

$$g(x) = \sum_{l \in \mathbb{Z}} \exp(-|x + l|),$$

vérifie l'hypothèse (5.4) avec  $\delta = 2$ .

Si on configure les estimateurs de seuillage par blocs avec les quantités définies au Théorème 5.6.4 alors on peut appliquer tous les résultats des sections précédentes.

### 5.6.3.4 Récapitulatif minimax concernant l'estimateur BlockShrink

Dans le cadre du modèle de déconvolution (5.3), le Théorème 5.6.5 synthétise les bornes supérieures de l'estimateur BlockShrink sous le risque  $\mathbb{L}^p$  sur des boules de Besov. C'est une conséquence des Théorèmes 5.4.1 et 5.4.2.

**THÉORÈME 5.6.5 (CONSÉQUENCE DES THÉORÈMES 5.4.1 ET 5.4.2)**

*Soit  $p \in [1, \infty[$ . Considérons le modèle de déconvolution défini par (5.3). Configurons l'estimateur BlockShrink  $\hat{f}_n^b$  défini par (5.1) avec  $l_j \asymp (\log n)^{(p/2) \vee 1}$ , les éléments décrits au Théorème 5.6.4 et la constante de seuillage  $\mu = \mu_1$ . Alors il existe une constante  $C > 0$  telle que, pour tout  $\pi \in [1, \infty]$   $s \in ]1/\pi, N]$ ,  $r \in [1, \infty]$  et  $n$  suffisamment grand, on a*

$$\sup_{f \in B_{\pi, r}^s(L)} \mathbb{E}_f^n(\|\hat{f}_n^b - f\|_p^p) \leq C\varphi_n,$$

où

$$\varphi_n = \begin{cases} n^{-\alpha_1 p} (\log n)^{\alpha_1 p 1_{\{p > \pi\}}}, & \text{si } \epsilon > 0, \\ (\log n/n)^{\alpha_2 p} (\log n)^{(p-\pi/r)+1_{\{\epsilon=0\}}}, & \text{si } \epsilon \leq 0, \end{cases}$$

avec  $\alpha_1 = s/(2(s + \delta) + 1)$ ,  $\alpha_2 = (s - 1/\pi + 1/p)/(2(s - 1/\pi + \delta) + 1)$  et  $\epsilon = \pi s + (\delta + 1/2)(\pi - p)$ .

L'estimateur  $\hat{f}_n^b$  est optimal, sauf pour le cas  $\epsilon > 0$  avec  $p > \pi$ .

**REMARQUE :** L'estimateur BlockShrink a de meilleures performances minimax que l'estimateur de seuillage dur développé par Johnstone et al. (2004).

## 5.6.4 Vers la pratique

Si on reprend pas à pas les preuves des Théorèmes 5.4.1 et 5.4.2 alors on observe que plus  $\mu_1$  est grand, plus la constante en facteur des vitesses de convergence l'est aussi. Comme on travaille avec un  $n$  aussi grand que l'on veut, cette constante est négligeable par rapport aux vitesses de convergence. Cependant, en pratique, l'entier  $n$  est fixé. La valeur de cette constante est donc à prendre en considération. Pour optimiser les performances de nos estimateurs de seuillage par blocs, il est donc préférable de prendre le plus petit  $\mu_1$  possible.

### 5.6.4.1 Choix de la constante de seuillage

Plaçons nous dans le cadre du modèle de bruit blanc gaussien (1.3). Lorsque  $p = 2$ , la constante de seuillage  $\mu_1$  considérée par Cai (1996; 1997) est le plus petit réel  $\mu$  vérifiant

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\sup_{b \in \{1, \dots, n(\log n)^{-1}\}} \sum_{i=(b-1)\log n}^{b\log n-1} |z_i|^2 \geq \mu \log n\right) = 0,$$

où  $(z_1, \dots, z_n)$  sont des variables aléatoires i.i.d de loi  $\mathcal{N}(0, 1)$ . Comme

$$\sum_{i=(b-1)\log n}^{b\log n-1} |z_i|^2 \sim \chi^2(\log n),$$

où  $\chi^2(\log n)$  désigne la loi du khi-deux à  $\log n$  degré de liberté, il est possible d'avoir une évaluation précise de celle-ci. Elle est solution de l'équation  $x - \log(x) = 3$ , d'où la valeur  $\mu_1 = 4,50524\dots$ . C'est avec cette constante que l'estimateur BlockShrink définit avec  $p = 2$  donne les meilleurs résultats numériques et visuels.

La question suivante s'est posée à nous : quelle est l'influence du facteur  $p$  dans les performances numériques et visuelles de la version  $\mathbb{L}^p$  de l'estimateur BlockShrink ?

À ce jour, il est difficile d'apporter une réponse définitive. En effet, pour  $p \geq 2$ , nous n'avons pas encore les outils mathématiques pour isoler le plus petit  $\mu > 0$  vérifiant

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \sup_{b \in \{0, \dots, n(\log n)^{-p/2}\}} \sum_{i=(b-1)(\log n)^{p/2}}^{b(\log n)^{p/2}-1} |z_i|^p \geq \mu(\log n)^{p/2} \right) = 0.$$

D'un point de vue théorique, la constante de seuillage idéale est la plus petite valeur réelle  $\mu$  indépendante de  $f$  telle que, pour tout  $j \in \{j_1, \dots, j_2\}$  et  $K \in \mathcal{A}_j$ , on a

$$\lim_{n \rightarrow \infty} \mathbb{P}_f^n \left( (l_j^{-1} \sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^p)^{1/p} \geq 2^{-1} \mu 2^{\delta j} n^{-1/2} \right) = 0.$$

**REMARQUE :** Lorsque  $p = 2$ , il existe des estimateurs en ondelettes de type seuillage par blocs ayant de meilleures performances numériques et visuelles que l'estimateur BlockShrink. Il y a notamment l'estimateur NeighBlock et l'estimateur NeighCoeff développés par Cai et Silverman (2001), lesquels sont des versions ondelettes de l'estimateur de Stein par blocs. Du point de vue théorique, ce dernier possède d'excellentes propriétés minimax, aussi bien au niveau des vitesses de convergence que des constantes qui les factorisent. Pour une étude complète, voir les articles de Cavalier et Tsybakov (2001) ainsi que le livre de Tsybakov (2004).

#### 5.6.4.2 Simulations

Ci-dessous, deux simulations graphiques comparant le rendu visuel de l'estimateur de seuillage global et de l'estimateur BlockShrink par rapport à l'estimateur de seuillage dur. Nous avons considéré le modèle de régression à pas équidistant (1.2),  $Y_i = f(i/n) + 0,04z_i$ . Nous avons pris la base Symlet[8],  $p = 2$ ,  $n = 2^{11}$ ,  $\mu = 0,04\sqrt{4,5}$  et les estimateurs empiriques des coefficients.

FIG. 5.1 – Simulation comparant le rendu graphique de l'estimateur de seuillage global et de l'estimateur de seuillage dur dans la reconstruction d'une fonction bruitée.

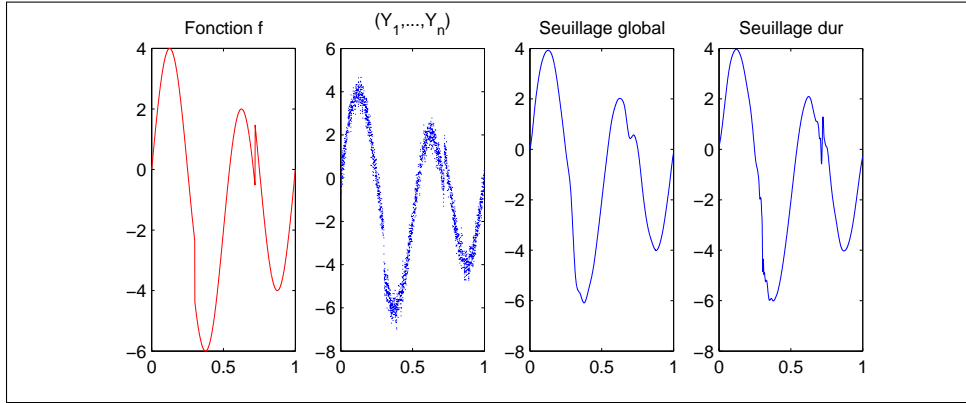
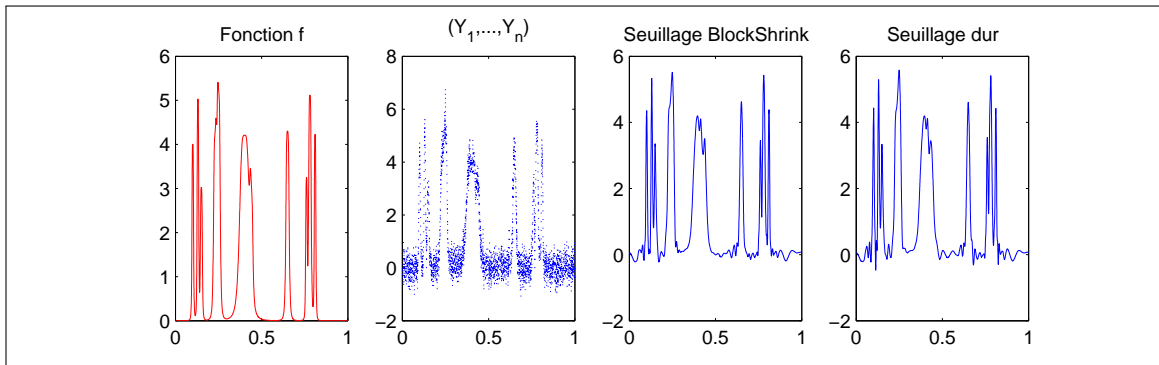


FIG. 5.2 – Simulation comparant le rendu visuel de l'estimateur BlockShrink et de l'estimateur de seuillage dur dans la reconstruction d'une fonction bruitée.



Ces figures sont représentatives de ce que l'on observe généralement en pratique : l'estimateur de seuillage dur a un meilleur rendu visuel que l'estimateur de seuillage global et l'estimateur BlockShrink a un meilleur rendu visuel que l'estimateur de seuillage dur.

## 5.7 Démonstrations

Les constantes  $c$  et  $C$  ne dépendent ni de  $f$ , ni de  $n$ . Elles peuvent prendre des valeurs différentes d'un terme à l'autre. On suppose que  $n$  est suffisamment grand.

### 5.7.1 Preuves des résultats de la Section 5.3

**Preuve du Théorème 5.3.1.** Décomposons la fonction  $f$  sur  $\zeta$  à partir du niveau  $j_1$ . La propriété d'inconditionnalité (2.3) implique

$$\begin{aligned} & \mathbb{E}_f^n(\|\hat{f}_n^* - f\|_p^p) \\ & \geq C(\mathbb{E}_f^n(\|(\sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\hat{\beta}_{j,k} 1_{\{\hat{b}_{j,K} \geq 2^{\delta j} \mu_1 n^{-1/2}\}} - \beta_{j,k}|^2 |\psi_{j,k}|^2 \\ & + \sum_{j=j_2+1}^{\infty} \sum_{k=0}^{2^j-1} |\beta_{j,k} \psi_{j,k}|^2)^{1/2}\|_p^p)) \\ & \geq C(T_1 \vee T_2), \end{aligned} \tag{5.5}$$

où

$$T_1 = \mathbb{E}_f^n(\|(\sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\beta_{j,k}|^2 1_{\{\hat{b}_{j,K} < 2^{\delta j} \mu_1 n^{-1/2}\}} |\psi_{j,k}|^2)^{1/2}\|_p^p)$$

et

$$T_2 = \|(\sum_{j=j_2+1}^{\infty} \sum_{k=0}^{2^j-1} |\beta_{j,k} \psi_{j,k}|^2)^{1/2}\|_p^p.$$

• *Majoration de  $\mathcal{Q}_2(f)$ .* La propriété d'inconditionnalité (2.3) et l'inégalité (5.5) entraînent

$$\mathcal{Q}_2(f) = \|\sum_{j=j_2+1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k}\|_p^p \leq CT_2 \leq C\mathbb{E}_f^n(\|\hat{f}_n^* - f\|_p^p). \tag{5.6}$$

• *Majoration de  $\mathcal{Q}_1(f)$ .* En appliquant l'inégalité de Minkowski et une inégalité élémentaire de convexité, on obtient

$$\mathcal{Q}_1(f) = \|\sum_{j=\tau}^{\infty} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{b_{j,K} \leq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} \psi_{j,k}\|_p^p \leq 3^{p-1}(W_1 + W_2 + W_3), \tag{5.7}$$

où

$$\begin{aligned} W_1 &= \|\sum_{j=\tau}^{j_1-1} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{b_{j,K} \leq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} \psi_{j,k}\|_p^p, \\ W_2 &= \|\sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{b_{j,K} \leq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} \psi_{j,k}\|_p^p \end{aligned}$$



et

$$W_3 = \left\| \sum_{j=j_2+1}^{\infty} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{b_{j,K} \leq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} \psi_{j,k} \right\|_p^p.$$

Étudions les majorations de  $W_1$ ,  $W_2$  et  $W_3$ .

• *Majoration de  $W_1$ .* En utilisant l'inégalité de Minkowski, une propriété élémentaire de la base d'ondelettes en norme  $\mathbb{L}^p$  (2.6), le fait que  $\text{Card}(\mathcal{A}_j) = 2^j l_j^{-1}$  et  $\alpha \in ]0, 1[$ , il vient

$$\begin{aligned} W_1 &\leq \left( \sum_{j=\tau}^{j_1-1} \left\| \sum_K \sum_{(K)} \beta_{j,k} \psi_{j,k} 1_{\{b_{j,K} \leq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} \right\|_p \right)^p \\ &\leq C \left( \sum_{j=\tau}^{j_1-1} 2^{j(1/2-1/p)} \left( \sum_K \sum_{(K)} |\beta_{j,k}|^p 1_{\{b_{j,K} \leq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} \right)^{1/p} \right)^p \\ &= C \left( \sum_{j=\tau}^{j_1-1} 2^{j(1/2-1/p)} l_j^{1/p} \left( \sum_K b_{j,K}^p 1_{\{b_{j,K} \leq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} \right)^{1/p} \right)^p \\ &\leq C n^{-p/2} \left( \sum_{j=\tau}^{j_1-1} 2^{j(\delta+1/2)} \right)^p \leq C n^{-p/2} 2^{j_1(\delta+1/2)p} \leq C n^{-\alpha p/2}. \end{aligned} \quad (5.8)$$

• *Majoration de  $W_2$ .* L'inégalité de Minkowski et une inégalité élémentaire de convexité entraînent

$$W_2 \leq 2^{p-1} (W_{2,1} + W_{2,2}),$$

où

$$W_{2,1} = \mathbb{E}_f^n \left( \left\| \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{b_{j,K} \leq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} 1_{\{\hat{b}_{j,K} < 2^{\delta j} \mu_1 n^{-1/2}\}} \psi_{j,k} \right\|_p^p \right)$$

et

$$W_{2,2} = \mathbb{E}_f^n \left( \left\| \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{b_{j,K} \leq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} 1_{\{\hat{b}_{j,K} \geq 2^{\delta j} \mu_1 n^{-1/2}\}} \psi_{j,k} \right\|_p^p \right).$$

– *Majoration de  $W_{2,1}$ .* La propriété d'inconditionnalité (2.3) et l'inégalité (5.5) entraînent

$$\begin{aligned} W_{2,1} &\leq C \mathbb{E}_f^n \left( \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\beta_{j,k}|^2 1_{\{\hat{b}_{j,K} < 2^{\delta j} \mu_1 n^{-1/2}\}} |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \right) \\ &= C T_1 \leq C \mathbb{E}_f^n (\|\hat{f}_n^* - f\|_p^p). \end{aligned}$$

– *Majoration de  $W_{2,2}$ .* En vertu de l'inégalité de Minkowski  $l_p$ , on a

$$\begin{aligned} 1_{\{b_{j,K} \leq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} 1_{\{\hat{b}_{j,K} \geq 2^{\delta j} \mu_1 n^{-1/2}\}} &\leq 1_{\{|\hat{b}_{j,K} - b_{j,K}| \geq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} \\ &\leq 1_{\{(l_j^{-1} \sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^p)^{1/p} \geq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}}. \end{aligned} \quad (5.9)$$

## 5.7. DÉMONSTRATIONS

---

Étudions séparément le cas où  $1 < p \leq 2$  et le cas où  $p > 2$ .

– Si  $1 < p \leq 2$ . En utilisant la propriété d'inconditionnalité (2.3), l'inégalité de Jensen, l'inégalité (5.9), l'hypothèse (F2) et de nouveau (2.3), il vient

$$\begin{aligned}
W_{2,2} &\leq C \mathbb{E}_f^n \left( \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\beta_{j,k}|^2 1_{\{b_{j,K} \leq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} 1_{\{\hat{b}_{j,K} \geq 2^{\delta j} \mu_1 n^{-1/2}\}} |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \right) \\
&\leq C \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\beta_{j,k}|^2 \mathbb{E}_f^n (1_{\{b_{j,K} \leq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} 1_{\{\hat{b}_{j,K} \geq 2^{\delta j} \mu_1 n^{-1/2}\}}) |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \\
&\leq C \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\beta_{j,k}|^2 \mathbb{P}_f^n \left( (l_j^{-1} \sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^p)^{1/p} \geq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2} \right) |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \\
&\leq C n^{-p/2} \left\| \left( \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^2 |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \leq C \|f\|_p^p n^{-p/2} \leq C n^{-\alpha p/2}.
\end{aligned}$$

– Si  $p > 2$ . Par la propriété d'inconditionnalité (2.3), l'inégalité de Minkowski généralisée, l'inégalité (5.9), l'hypothèse (F2) et de nouveau (2.3), il vient

$$\begin{aligned}
W_{2,2} &\leq C \mathbb{E}_f^n \left( \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\beta_{j,k}|^2 1_{\{b_{j,K} \leq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} 1_{\{\hat{b}_{j,K} \geq 2^{\delta j} \mu_1 n^{-1/2}\}} |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \right) \\
&\leq C \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\beta_{j,k}|^2 \mathbb{E}_f^n (1_{\{b_{j,K} \leq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} 1_{\{\hat{b}_{j,K} \geq 2^{\delta j} \mu_1 n^{-1/2}\}})^{2/p} |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \\
&\leq C \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\beta_{j,k}|^2 \mathbb{P}_f^n \left( (l_j^{-1} \sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^p)^{1/p} \geq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2} \right)^{2/p} |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \\
&\leq C n^{-p/2} \left\| \left( \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^2 |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \leq C \|f\|_p^p n^{-p/2} \leq C n^{-\alpha p/2}.
\end{aligned}$$

En mettant les majorations de  $W_{2,1}$  et  $W_{2,2}$  ensemble, pour tout  $p > 1$ , on a

$$W_2 \leq C n^{-\alpha p/2}. \quad (5.10)$$

• *Majoration de  $W_3$ .* La propriété d'inconditionnalité (2.3) entraîne

$$W_3 \leq C \left\| \sum_{j=j_2+1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k} \right\|_p^p \leq C T_2 \leq C \mathbb{E}_f^n (\|\hat{f}_n^* - f\|_p^p). \quad (5.11)$$

En mettant (5.7), (5.8), (5.10) et (5.11) bout à bout, pour tout  $p > 1$ , on obtient

$$\mathcal{Q}_1(f) \leq C (\mathbb{E}_f^n (\|\hat{f}_n^* - f\|_p^p) + n^{-\alpha p/2}).$$

On en déduit l'existence d'une constante  $C > 0$  telle que

$$\mathcal{Q}_1(f) + \mathcal{Q}_2(f) \leq C(\mathbb{E}_f^n(\|\hat{f}_n^* - f\|_p^p) + n^{-\alpha p/2}).$$

Cela achève la preuve du Théorème 5.3.1.  $\square$

**Preuve du Théorème 5.3.2.** On suppose que  $p > 1$ . Lorsque  $p = 1$ , le résultat souhaité s'obtient sans propriété particulière concernant la géométrie des bases d'ondelettes.

Décomposons la fonction  $f$  sur  $\zeta$  à partir du niveau de résolution  $j_1$ . En combinant l'inégalité de Minkowski avec une inégalité élémentaire de convexité, on a

$$\mathbb{E}_f^n(\|\hat{f}_n^* - f\|_p^p) \leq 4^{p-1}(G_1 + G_2 + G_3 + \mathcal{Q}_2(f)), \quad (5.12)$$

où

$$G_1 = \mathbb{E}_f^n(\|\sum_{k=0}^{2^{j_1}-1} (\hat{\alpha}_{j_1,k} - \alpha_{j_1,k}) \phi_{j_1,k}\|_p^p),$$

$$G_2 = \mathbb{E}_f^n(\|\sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{\hat{b}_{j,K} < 2^{\delta j} \mu_1 n^{-1/2}\}} \psi_{j,k}\|_p^p)$$

et

$$G_3 = \mathbb{E}_f^n(\|\sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} (\hat{\beta}_{j,k} - \beta_{j,k}) 1_{\{\hat{b}_{j,K} \geq 2^{\delta j} \mu_1 n^{-1/2}\}} \psi_{j,k}\|_p^p).$$

Analysons les majorations des termes  $G_1$ ,  $G_2$  et  $G_3$ .

- *Majoration de  $G_1$ .* Il découle de une propriété élémentaire de la base d'ondelettes en norme  $\mathbb{L}^p$  (2.6) et de l'hypothèse (F1) que

$$\begin{aligned} G_1 &\leq C 2^{j_1(p/2-1)} \sum_{k=0}^{2^{j_1}-1} \mathbb{E}_f^n(|\hat{\alpha}_{j_1,k} - \alpha_{j_1,k}|^p) \leq C n^{-p/2} 2^{j_1(\delta+1/2)p} \\ &\leq C n^{-p/2} (\log n)^{((p/2) \vee 1)(\delta+1/2)p} \leq C n^{-\alpha p/2}. \end{aligned} \quad (5.13)$$

- *Majoration de  $G_2$ .* En vertu de l'inégalité de Minkowski et d'une inégalité élémentaire de convexité, on a

$$G_2 \leq 2^{p-1}(G_{2,1} + G_{2,2}),$$

où

$$G_{2,1} = \mathbb{E}_f^n(\|\sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{\hat{b}_{j,K} < 2^{\delta j} \mu_1 n^{-1/2}\}} 1_{\{b_{j,K} \leq 2^{2\delta j} \mu_1 n^{-1/2}\}} \psi_{j,k}\|_p^p)$$

et

$$G_{2,2} = \mathbb{E}_f^n(\|\sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{\hat{b}_{j,K} < 2^{\delta j} \mu_1 n^{-1/2}\}} 1_{\{b_{j,K} > 2^{2\delta j} \mu_1 n^{-1/2}\}} \psi_{j,k}\|_p^p).$$

## 5.7. DÉMONSTRATIONS

---

– *Majoration de  $G_{2,1}$ .* Étudions séparément le cas où  $1 < p \leq 2$  et le cas où  $p > 2$ .

– *Si  $1 < p \leq 2$ .* Par la propriété d'inconditionnalité (2.3) et une inégalité de norme  $l_p$ , il vient

$$\begin{aligned}
 G_{2,1} &\leq C \left\| \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{b_{j,K} \leq 2^{2\delta j} \mu_1 n^{-1/2}\}} \psi_{j,k} \right\|_p^p \\
 &\leq C \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\beta_{j,k}|^2 1_{\{b_{j,K} \leq 2^{2\delta j} \mu_1 n^{-1/2}\}} |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \\
 &\leq C \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\beta_{j,k}|^p 1_{\{b_{j,K} \leq 2^{2\delta j} \mu_1 n^{-1/2}\}} 2^{j(p/2-1)} \\
 &= C \left\| \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{b_{j,K} \leq 2^{2\delta j} \mu_1 n^{-1/2}\}} \psi_{j,k} \right\|_{B_{p,p}^0}^p \leq C \mathcal{Q}_1^*(f).
 \end{aligned}$$

– *Si  $p > 2$ .* Par la propriété d'inconditionnalité (2.3), il vient

$$G_{2,1} \leq C \left\| \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{b_{j,K} \leq 2^{2\delta j} \mu_1 n^{-1/2}\}} \psi_{j,k} \right\|_p^p \leq C \mathcal{Q}_1^*(f).$$

– *Majoration de  $G_{2,2}$ .* Remarquons que l'inégalité de Minkowski  $l_p$  implique

$$\begin{aligned}
 1_{\{b_{j,K} > 2^{2\delta j} \mu_1 n^{-1/2}\}} 1_{\{\hat{b}_{j,K} < 2^{\delta j} \mu_1 n^{-1/2}\}} &\leq 1_{\{|\hat{b}_{j,K} - b_{j,K}| \geq 2^{\delta j} \mu_1 n^{-1/2}\}} \\
 &\leq 1_{\{(l_j^{-1} \sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^p)^{1/p} \geq 2^{\delta j} \mu_1 n^{-1/2}\}}.
 \end{aligned} \tag{5.14}$$

Distinguons le cas où  $1 < p \leq 2$  et le cas où  $p > 2$ .

– *Si  $1 < p \leq 2$ .* Par la propriété d'inconditionnalité (2.3), l'inégalité de Jensen, l'inégalité (5.14), une inégalité de norme  $l_p$ , l'hypothèse (F2) et de nouveau (2.3), on a

$$\begin{aligned}
 G_{2,2} &\leq C \mathbb{E}_f^n \left( \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\beta_{j,k}|^2 1_{\{b_{j,K} > 2^{2\delta j} \mu_1 n^{-1/2}\}} 1_{\{\hat{b}_{j,K} < 2^{\delta j} \mu_1 n^{-1/2}\}} |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \right) \\
 &\leq C \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\beta_{j,k}|^2 \mathbb{E}_f^n (1_{\{b_{j,K} > 2^{2\delta j} \mu_1 n^{-1/2}\}} 1_{\{\hat{b}_{j,K} < 2^{\delta j} \mu_1 n^{-1/2}\}}) |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \\
 &\leq C \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\beta_{j,k}|^2 \mathbb{P}_f^n \left( (l_j^{-1} \sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^p)^{1/p} \geq 2^{\delta j} \mu_1 n^{-1/2} \right) |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \\
 &\leq C n^{-p/2} \left\| \left( \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^2 |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \leq C \|f\|_p^p n^{-p/2} \leq C n^{-\alpha p/2}.
 \end{aligned}$$

– Si  $p > 2$ . En utilisant la propriété d'inconditionnalité (2.3), l'inégalité de Minkowski généralisée, l'inégalité (5.14), l'hypothèse (F2) et de nouveau (2.3), on trouve

$$\begin{aligned}
 G_{2,2} &\leq C \mathbb{E}_f^n \left( \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\beta_{j,k}|^2 1_{\{b_{j,K} > 2^{2\delta j} \mu_1 n^{-1/2}\}} 1_{\{\hat{b}_{j,K} < 2^{\delta j} \mu_1 n^{-1/2}\}} |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \right) \\
 &\leq C \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\beta_{j,k}|^2 \mathbb{E}_f^n (1_{\{b_{j,K} > 2^{2\delta j} \mu_1 n^{-1/2}\}} 1_{\{\hat{b}_{j,K} < 2^{\delta j} \mu_1 n^{-1/2}\}})^{2/p} |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \\
 &\leq C \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\beta_{j,k}|^2 \mathbb{P}_f^n \left( (l_j^{-1} \sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^p)^{1/p} \geq 2^{\delta j} \mu_1 n^{-1/2} \right)^{2/p} |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \\
 &\leq C n^{-p/2} \left\| \left( \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^2 |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \leq C \|f\|_p^p n^{-\alpha p/2} \leq C n^{-\alpha p/2}.
 \end{aligned}$$

Il découle des majorations obtenues que

$$G_2 \leq C(\mathcal{Q}_1^*(f) + n^{-\alpha p/2}). \quad (5.15)$$

• *Majoration de  $G_3$ .* L'inégalité de Cauchy-Schwartz combinée avec les hypothèses (F1) et (F2) donne

$$\begin{aligned}
 &\mathbb{E}_f^n (|\hat{\beta}_{j,k} - \beta_{j,k}|^p 1_{\{\hat{b}_{j,K} \geq 2^{\delta j} \mu_1 n^{-1/2}\}} 1_{\{b_{j,K} < 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}}) \\
 &\leq \mathbb{E}_f^n (|\hat{\beta}_{j,k} - \beta_{j,k}|^{2p})^{1/2} \mathbb{P}_f^n ((l_j^{-1} \sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^p)^{1/p} \geq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2})^{1/2} \\
 &\leq C 2^{\delta j p} n^{-p}.
 \end{aligned} \quad (5.16)$$

Par l'inégalité de Minkowski avec une inégalité de norme  $l_p$ , il vient

$$G_3 \leq 2^{p-1} (G_{3,1} + G_{3,2}),$$

où

$$G_{3,1} = \mathbb{E}_f^n \left( \left\| \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} (\hat{\beta}_{j,k} - \beta_{j,k}) 1_{\{\hat{b}_{j,K} \geq 2^{\delta j} \mu_1 n^{-1/2}\}} 1_{\{b_{j,K} < 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} \psi_{j,k} \right\|_p^p \right)$$

et

$$G_{3,2} = \mathbb{E}_f^n \left( \left\| \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} (\hat{\beta}_{j,k} - \beta_{j,k}) 1_{\{\hat{b}_{j,K} \geq 2^{\delta j} \mu_1 n^{-1/2}\}} 1_{\{b_{j,K} \geq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} \psi_{j,k} \right\|_p^p \right).$$

– *Majoration de  $G_{3,1}$ .* Distinguons le cas où  $1 < p \leq 2$  et le cas où  $p > 2$ .

## 5.7. DÉMONSTRATIONS

---

– Si  $1 < p \leq 2$ . Par la propriété d'inconditionnalité (2.3), une inégalité de norme  $l_p$ , l'inégalité (5.16) et le fait que  $\nu \in ]0, (\delta + 1)^{-1}]$ , on a

$$\begin{aligned}
G_{3,1} &\leq C \mathbb{E}_f^n \left( \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^2 1_{\{\hat{b}_{j,K} \geq 2^{\delta j} \mu_1 n^{-1/2}\}} 1_{\{b_{j,K} < 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \right) \\
&\leq C \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \mathbb{E}_f^n (|\hat{\beta}_{j,k} - \beta_{j,k}|^p 1_{\{\hat{b}_{j,K} \geq 2^{\delta j} \mu_1 n^{-1/2}\}} 1_{\{b_{j,K} < 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}}) \|\psi_{j,k}\|_p^p \\
&\leq C n^{-p} \sum_{j=\tau}^{j_2} \sum_{k=0}^{2^j-1} 2^{\delta j p} \|\psi_{j,k}\|_p^p = C n^{-p} \sum_{j=\tau}^{j_2} 2^{j(\delta+1/2)p} \leq C n^{-p} 2^{j_2(\delta+1/2)p} \\
&\leq C n^{-p} n^{\nu p(\delta+1/2)} \leq C n^{-\alpha p/2}.
\end{aligned}$$

– Si  $p > 2$ . Par la propriété d'inconditionnalité (2.3), l'inégalité de Minskowski généralisée, l'inégalité (5.9), la propriété de Temlyakov (2.4) et des inégalités similaires au cas  $1 < p < 2$ , il vient

$$\begin{aligned}
G_{3,1} &\leq C \mathbb{E}_f^n \left( \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^2 1_{\{\hat{b}_{j,K} \geq 2^{\delta j} \mu_1 n^{-1/2}\}} 1_{\{b_{j,K} < 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \right) \\
&\leq C \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \mathbb{E}_f^n (|\hat{\beta}_{j,k} - \beta_{j,k}|^p 1_{\{\hat{b}_{j,K} \geq 2^{\delta j} \mu_1 n^{-1/2}\}} 1_{\{b_{j,K} < 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}}) \right)^{2/p} |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \\
&\leq C n^{-p} \left\| \left( \sum_{j=\tau}^{j_2} \sum_{k=0}^{2^j-1} 2^{2\delta j} |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \leq C n^{-p} \sum_{j=\tau}^{j_2} \sum_{k=0}^{2^j-1} 2^{\delta j p} \|\psi_{j,k}\|_p^p \leq C n^{-\alpha p/2}.
\end{aligned}$$

– *Majoration de  $G_{3,2}$ .* Étudions séparément le cas où  $1 < p \leq 2$  et le cas où  $p > 2$ .

– Si  $1 < p \leq 2$ . En utilisant la propriété d'inconditionnalité (2.3), une inégalité de norme  $l_p$  et l'hypothèse (F1), on trouve

$$\begin{aligned}
G_{3,2} &\leq C \mathbb{E}_f^n \left( \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^2 1_{\{b_{j,K} \geq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \right) \\
&\leq C \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \mathbb{E}_f^n (|\hat{\beta}_{j,k} - \beta_{j,k}|^p 1_{\{b_{j,K} \geq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}}) \|\psi_{j,k}\|_p^p \\
&\leq C n^{-p/2} \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} 1_{\{b_{j,K} \geq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} 2^{\delta j p} \|\psi_{j,k}\|_p^p = C G_{3,2}^*
\end{aligned}$$

Par un découpage adéquat et l'inégalité de Markov, on a

$$\begin{aligned}
 G_{3,2}^* &= n^{-p/2} \sum_{m \in \mathbb{N}} \sum_{j=j_1}^{j_2} \sum_K 1_{\{2^{-1}\mu_1 n^{-1/2} 2^{\delta j} 2^m \leq b_{j,K} < 2^{\delta j} 2^{-1}\mu_1 n^{-1/2} 2^{m+1}\}} 2^{\delta j p} l_j 2^{j(p/2-1)} \\
 &\leq C \sum_{m \in \mathbb{N}} 2^{-mp} \sum_{j=j_1}^{j_2} \sum_K (b_{j,K})^p 1_{\{b_{j,K} < 2^{\delta j} 2^{-1}\mu_1 n^{-1/2} 2^{m+1}\}} l_j 2^{j(p/2-1)} \\
 &= C \sum_{m \in \mathbb{N}} 2^{-mp} \left\| \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{b_{j,K} < 2^{\delta j} 2^{-1}\mu_1 n^{-1/2} 2^{m+1}\}} \psi_{j,k} \right\|_{B_{p,p}^0}^p = C \mathcal{Q}_1^*(f).
 \end{aligned}$$

– Si  $p > 2$ . En utilisant la propriété d'inconditionnalité (2.3), l'inégalité de Minkowski généralisée et l'hypothèse (F1), on obtient

$$\begin{aligned}
 G_{3,2} &\leq C \mathbb{E}_f^n \left( \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^2 1_{\{b_{j,K} \geq 2^{\delta j} 2^{-1}\mu_1 n^{-1/2}\}} |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \right) \\
 &\leq C \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \mathbb{E}_f^n (|\hat{\beta}_{j,k} - \beta_{j,k}|^p) 1_{\{b_{j,K} \geq 2^{\delta j} 2^{-1}\mu_1 n^{-1/2}\}} |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p \\
 &\leq C n^{-p/2} \left\| \left( \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} 1_{\{b_{j,K} \geq 2^{\delta j} 2^{-1}\mu_1 n^{-1/2}\}} 2^{2\delta j} |\psi_{j,k}|^2 \right)^{1/2} \right\|_p^p = C G_{3,2}^*.
 \end{aligned}$$

Par la propriété de Temlyakov (2.4), des techniques de majoration identiques au cas où  $1 < p < 2$  et l'inclusion  $B_{p,p}^0 \subseteq \mathbb{L}^p$ , on trouve

$$\begin{aligned}
 G_{3,2}^* &\leq C n^{-p/2} \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} 1_{\{b_{j,K} \geq 2^{\delta j} 2^{-1}\mu_1 n^{-1/2}\}} 2^{\delta j p} \|\psi_{j,k}\|_p^p \\
 &\leq C \sum_{m \in \mathbb{N}} 2^{-mp} \left\| \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{b_{j,K} < 2^{\delta j} 2^{-1}\mu_1 n^{-1/2} 2^{m+1}\}} \psi_{j,k} \right\|_{B_{p,p}^0}^p \\
 &\leq C \sum_{m \in \mathbb{N}} 2^{-mp} \left\| \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{b_{j,K} < 2^{\delta j} 2^{-1}\mu_1 n^{-1/2} 2^{m+1}\}} \psi_{j,k} \right\|_p^p = C \mathcal{Q}_1^*(f).
 \end{aligned} \tag{5.17}$$

Pour tout  $p > 1$ , les majorations obtenues impliquent

$$G_3 \leq C(\mathcal{Q}_1^*(f) + n^{-\alpha p/2}). \tag{5.18}$$

En mettant les inégalités (5.12), (5.13), (5.15) et (5.18) ensemble, on a

$$\mathbb{E}_f^n(\|\hat{f}_n^* - f\|_p^p) \leq C(\mathcal{Q}_1^*(f) + \mathcal{Q}_2(f) + n^{-\alpha p/2}).$$

La preuve du Théorème 5.3.2 est complète.  $\square$

### 5.7.2 Preuves des résultats de la Section 5.4

**Preuve du Théorème 5.4.1.** Soit  $\pi \geq p$ . On va uniquement traiter le cas où  $p \geq 2$ . Le cas où  $1 \leq p < 2$  se prouvant exactement de la même façon.

Par le Théorème 5.3.2, il suffit de montrer que pour tout  $f \in B_{\pi,r}^s(L)$  il existe une constante  $C > 0$  vérifiant

$$\mathcal{Q}_1^*(f) \vee \mathcal{Q}_2(f) \leq Cn^{-\alpha_1 p}, \quad \alpha_1 = s/(2(s+\delta)+1).$$

• *Majoration de  $\mathcal{Q}_1^*(f)$ .* Pour tout  $m \in \mathbb{N}$ , considérons un entier  $j_3$  tel que

$$2^{j_3} \asymp 2^{-m/(2s)} n^{1/(2(s+\delta)+1)}.$$

Par l'inégalité de Minkowski, une inégalité élémentaire de convexité et la propriété d'inconditionnalité (2.3), il vient

$$\mathcal{Q}_1^*(f) = \sum_{m \in \mathbb{N}} 2^{-mp} \left\| \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{b_{j,K} \leq \mu_1 2^{\delta j} 2^{m n^{-1/2}}\}} \psi_{j,k} \right\|_p^p \leq 2^{p-1} (S_1 + S_2),$$

où

$$S_1 = \sum_{m \in \mathbb{N}} 2^{-mp} \left\| \sum_{j=\tau}^{j_3} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{b_{j,K} \leq \mu_1 2^{\delta j} 2^{m n^{-1/2}}\}} \psi_{j,k} \right\|_p^p$$

et

$$S_2 = \sum_{m \in \mathbb{N}} 2^{-mp} \left\| \sum_{j=j_3+1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k} \right\|_p^p.$$

Étudions les majorations de  $S_1$  et  $S_2$ .

– *Majoration de  $S_1$ .* Si  $b_{j,K} \leq \mu_1 2^{\delta j} 2^{m n^{-1/2}}$  alors on a

$$\left( \sum_{(K)} |\beta_{j,k}|^p \right)^{1/p} \leq \mu_1 n^{-1/2} 2^m 2^{\delta j} l_j^{1/p}.$$

Par l'inégalité de Minkowski et une propriété élémentaire de la base d'ondelettes en norme  $\mathbb{L}^p$  (2.6), il vient

$$\begin{aligned} S_1 &\leq C \sum_{m \in \mathbb{N}} 2^{-mp} \left( \sum_{j=\tau}^{j_3} 2^{j(1/2-1/p)} \left( \sum_K \sum_{(K)} |\beta_{j,k}|^p 1_{\{b_{j,K} \leq \mu_1 2^{\delta j} 2^{m n^{-1/2}}\}} \right)^{1/p} \right)^p \\ &\leq C n^{-p/2} \sum_{m \in \mathbb{N}} \left( \sum_{j=\tau}^{j_3} 2^{j(1/2-1/p)} (Card(\mathcal{A}_j) 2^{\delta j p} l_j)^{1/p} \right)^p \\ &\leq C n^{-p/2} \sum_{m \in \mathbb{N}} 2^{j_3(\delta+1/2)p} \leq C n^{-sp/(2(s+\delta)+1)} \sum_{m \in \mathbb{N}} 2^{-mp(1+2\delta)/(4s)} \leq C n^{-\alpha_1 p}. \end{aligned}$$



– *Majoration de  $S_2$* . En utilisant l'inégalité Minkowski, une propriété élémentaire de la base d'ondelettes en norme  $\mathbb{L}^p$  (2.6) et l'inclusion  $B_{\pi,r}^s(L) \subseteq B_{p,\infty}^s(L)$ , on trouve

$$\begin{aligned} S_2 &\leq C \sum_{m \in \mathbb{N}} 2^{-mp} \left( \sum_{j=j_3+1}^{\infty} 2^{j(1/2-1/p)} \left( \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \right)^{1/p} \right)^p \leq C \sum_{m \in \mathbb{N}} 2^{-mp} \left( \sum_{j=j_3+1}^{\infty} 2^{-js} \right)^p \\ &\leq C \sum_{m \in \mathbb{N}} 2^{-mp} 2^{-j_3 sp} \leq C n^{-sp/(2(s+\delta)+1)} \sum_{m \in \mathbb{N}} 2^{-mp/2} \leq C n^{-\alpha_1 p}. \end{aligned}$$

Les majorations précédentes justifient l'existence d'une constante  $C > 0$  telle que

$$\mathcal{Q}_1^*(f) \leq C n^{-\alpha_1 p}. \quad (5.19)$$

• *Majoration de  $\mathcal{Q}_2(f)$* . Par l'inégalité de Minkowski, une propriété élémentaire de la base d'ondelettes en norme  $\mathbb{L}^p$  (2.6), l'inclusion  $B_{\pi,r}^s(L) \subseteq B_{p,r}^s(L)$  et le fait que  $s > 1/\pi - \delta - 1/2 + 1/(2\nu)$ , on a

$$\begin{aligned} \mathcal{Q}_2(f) &\leq C \left( \sum_{j=j_2+1}^{\infty} 2^{j(1/2-1/p)} \left( \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \right)^{1/p} \right)^p \leq C \left( \sum_{j=j_2+1}^{\infty} 2^{-js} \right)^p \leq C 2^{-j_2 sp} \\ &\leq C (\log n/n)^{\nu sp} \leq C n^{-\alpha_1 p}. \end{aligned} \quad (5.20)$$

En combinant les inégalités (5.19) et (5.20) et en prenant  $\alpha = 2\alpha_1$ , le Théorème 5.3.2 nous permet de conclure.  $\square$

**Preuve de la Proposition 5.4.1.** Soit  $p > \pi$ . L'objectif est d'appliquer le Théorème 5.3.1. Soit  $j_3$  un entier tel que

$$c_0 n^{1/(2(s-1/\pi+1/p+\delta)+1)} \leq 2^{j_3} \leq C_0 n^{1/(2(s-1/\pi+1/p+\delta)+1)},$$

où  $c_0 > 0$  et  $C_0 > 0$  sont deux constantes choisies a posteriori. Considérons la fonction  $f_* : [0, 1] \rightarrow \mathbb{R}$  définie par

$$f_*(x) = L_* 2^{-j_3(s-1/\pi+1/2)} \psi_{j_3, \varrho}(x),$$

où  $L_* > 0$  est une constante choisie a posteriori et  $\varrho$  est un entier fixé dans  $\{0, \dots, 2^{j_3} - 1\}$ . Les coefficients d'ondelettes de  $f_*$  valent

$$\beta_{j,k} = \int_0^1 f_*(x) \psi_{j,k}(x) dx = \begin{cases} L_* 2^{-j_3(s-1/\pi+1/2)} & \text{si } j = j_3 \text{ et } k = \varrho, \\ 0 & \text{sinon.} \end{cases}$$

Par conséquent, on a  $2^{j_3(s+1/2)} (2^{-j_3} |\beta_{j_3, \varrho}|^\pi)^{1/\pi} = L_*$ . Par un choix convenable de  $L_*$ , on a donc  $f_* \in B_{\pi,r}^s(L)$ .

Pour  $\alpha \in ]0, 1[$ , le Théorème 5.3.1 entraîne

$$\mathcal{Q}_1(f_*) \leq C (\mathbb{E}_f^n(\|\hat{f}_n^g - f_*\|_p^p) + n^{-\alpha p/2}), \quad (5.21)$$

où

$$\mathcal{Q}_1(f_*) = \|\beta_{j_3, \varrho} 1_{\{b_{j_3} \leq 2^{\delta j_3} 2^{-1} \mu_1 n^{-1/2}\}} \psi_{j_3, \varrho}\|_p^p.$$

Par définition de  $j_3$ , en prenant  $c_0$  suffisamment grand, on a

$$\begin{aligned} b_{j_3} &= (2^{-j_3} |\beta_{j_3, \varrho}|^p)^{1/p} = L_* 2^{-j_3(s-1/\pi+1/p+1/2)} \leq L_* c_0^{-(s-1/\pi+1/p+1/2+\delta)} 2^{\delta j_3} \\ &\leq 2^{-1} \mu_1 n^{-1/2} 2^{\delta j_3}. \end{aligned}$$

Une propriété élémentaire de la base d'ondelettes en norme  $\mathbb{L}^p$  (2.6) entraîne

$$\begin{aligned} \mathcal{Q}_1(f_*) &= L^p 2^{-j_3(s-1/\pi+1/2)p} \|\psi_{j_3, \varrho}\|_p^p \geq c 2^{-j_3(s-1/\pi+1/2)p} 2^{j_3 p/2-1} = c 2^{-j_3(s-1/\pi+1/p)p} \\ &\geq c C_0^{-(s-1/\pi+1/p)p} n^{-\alpha_3 p}, \end{aligned} \quad (5.22)$$

où

$$\alpha_3 = (s - 1/\pi + 1/p)/(2(s - 1/\pi + 1/p + \delta) + 1).$$

En mettant les inégalités (5.21) et (5.22) ensemble et en prenant  $\alpha \in ]2\alpha_3, 1[$ , on prouve l'existence d'une constante  $c > 0$  telle que

$$n^{\alpha_3 p} \sup_{f \in B_{\pi, r}^s(L)} \mathbb{E}_f^n(\|\hat{f}_n^g - f\|_p^p) \geq n^{\alpha_3 p} \mathbb{E}_f^n(\|\hat{f}_n^g - f_*\|_p^p) \geq c.$$

Cela termine la preuve de la Proposition 5.4.1. □

**Preuve du Théorème 5.4.2.** Soit  $p \geq 2$ . Par le Théorème 5.3.2, il suffit de montrer que pour tout  $f \in B_{\pi, r}^s(L)$  il existe une constante  $C > 0$  vérifiant

$$\mathcal{Q}_1^*(f) \vee \mathcal{Q}_2(f) \leq \begin{cases} C (\log n/n)^{\alpha_1 p}, & \text{si } \epsilon > 0 \text{ et } p > \pi, \\ C (\log n/n)^{\alpha_2 p} (\log n)^{(p-\pi/r)+1_{\{\epsilon=0\}}}, & \text{si } \epsilon \leq 0, \end{cases}$$

avec  $\alpha_1 = s/(2(s + \delta) + 1)$ ,  $\alpha_2 = (s - 1/\pi + 1/p)/(2(s - 1/\pi + \delta) + 1)$  et  $\epsilon = \pi s + (\delta + 1/2)(\pi - p)$ .

• *Majoration de  $\mathcal{Q}_1^*(f)$ .* Considérons  $j_4$  un entier défini par

$$2^{j_4} \asymp \begin{cases} 2^{-m/(2s)} (n/\log n)^{1/(2(s+\delta)+1)}, & \text{si } \epsilon > 0 \text{ et } p > \pi, \\ 2^{-m/(2s)} (n/\log n)^{1/(2(s-1/\pi+\delta)+1)}, & \text{si } \epsilon \leq 0. \end{cases}$$

L'inégalité de Minkowski et une inégalité élémentaire de convexité entraînent :

$$\mathcal{Q}_1^*(f) \leq 2^{p-1}(T_1 + T_2),$$

où

$$T_1 = \sum_{m \in \mathbb{N}} 2^{-mp} \left\| \sum_{j=\tau}^{j_4} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{b_{j,K} \leq \mu_1 2^{\delta j} 2^{m n^{-1/2}}\}} \psi_{j,k} \right\|_p^p$$

et

$$T_2 = \sum_{m \in \mathbb{N}} 2^{-mp} \left\| \sum_{j=j_4+1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{b_{j,K} \leq \mu_1 2^{\delta j} 2^m n^{-1/2}\}} \psi_{j,k} \right\|_p^p.$$

Distinguons le cas où  $\epsilon > 0$  avec  $p > \pi$  et le cas où  $\epsilon \leq 0$ .

• Si  $\epsilon > 0$  avec  $p > \pi$ .

– Majoration de  $T_1$ . Si  $b_{j,K} \leq \mu_1 2^{\delta j} 2^m n^{-1/2}$  alors on a

$$\left( \sum_{(K)} |\beta_{j,k}|^p \right)^{1/p} \leq \mu_1 n^{-1/2} 2^m 2^{\delta j} l_j^{1/p}.$$

L'inégalité Minkowski et une propriété élémentaire de la base d'ondelettes en norme  $\mathbb{L}^p$  (2.6) entraînent

$$\begin{aligned} T_1 &\leq C \sum_{m \in \mathbb{N}} 2^{-mp} \left( \sum_{j=\tau}^{j_4} 2^{j(1/2-1/p)} \left( \sum_K \sum_{(K)} |\beta_{j,k}|^p 1_{\{b_{j,K} \leq \mu_1 2^{\delta j} 2^m n^{-1/2}\}} \right)^{1/p} \right)^p \\ &\leq C \sum_{m \in \mathbb{N}} n^{-p/2} \left( \sum_{j=\tau}^{j_4} 2^{j(1/2-1/p)} (Card(\mathcal{A}_j) 2^{\delta j p} (\log n)^{p/2})^{1/p} \right)^p \\ &\leq C n^{-p/2} \sum_{m \in \mathbb{N}} \left( \sum_{j=\tau}^{j_4} 2^{j(1/2+\delta)} \right)^p \leq C n^{-p/2} \sum_{m \in \mathbb{N}} 2^{j_4(1/2+\delta)p} \\ &\leq C (\log n/n)^{sp/(2(s+\delta)+1)} \sum_{m \in \mathbb{N}} 2^{-mp(1+2\delta)/(4s)} \leq C (\log n/n)^{\alpha_1 p}. \end{aligned}$$

– Majoration de  $T_2$ . Puisque  $l_j \asymp (\log n)^{p/2}$ , pour tout  $k$  dans  $\mathcal{B}_{j,K}$  il existe une constante  $C > 0$  vérifiant

$$\{b_{j,K} \leq \mu_1 2^{m+1} n^{-1/2} 2^{\delta j}\} \subseteq \{|\beta_{j,k}| \leq C \mu_1 2^{m+1} 2^{\delta j} \sqrt{(\log n/n)}\}. \quad (5.23)$$

## 5.7. DÉMONSTRATIONS

---

L'inclusion  $B_{\pi,r}^s(L) \subseteq B_{p,r}^{s-1/\pi+1/p}(L)$  et le fait que  $\epsilon > 0$  avec  $p > \pi$  entraînent

$$\begin{aligned}
T_2 &\leq C \sum_{m \in \mathbb{N}} 2^{-mp} \left( \sum_{j=j_4+1}^{j_2} 2^{j(1/2-1/p)} \left( \sum_K \sum_{(K)} |\beta_{j,k}|^p 1_{\{b_{j,K} \leq \mu_1 2^{\delta j} 2^m n^{-1/2}\}} \right)^{1/p} \right)^p \\
&\leq C (\log n)^{(p-\pi)/2} n^{(\pi-p)/2} \sum_{m \in \mathbb{N}} 2^{-m\pi} \left( \sum_{j=j_4+1}^{j_2} 2^{j(1/2-1/p)} 2^{\delta j((p-\pi)/p)} \left( \sum_{k=0}^{2^j-1} |\beta_{j,k}|^\pi \right)^{1/p} \right)^p \\
&\leq C (\log n)^{(p-\pi)/2} n^{(\pi-p)/2} \sum_{m \in \mathbb{N}} 2^{-m\pi} \left( \sum_{j=j_4+1}^{j_2} 2^{-j\epsilon/p} \right)^p \\
&\leq C (\log n)^{(p-\pi)/2} n^{(\pi-p)/2} \sum_{m \in \mathbb{N}} 2^{-m\pi} 2^{-j_4\epsilon} \\
&\leq C (\log n)^{(p-\pi)/2} n^{(\pi-p)/2} (\log n/n)^{\epsilon/(2(s+\delta)+1)} \sum_{m \in \mathbb{N}} 2^{-m\pi/2+m(2\delta+1)(\pi-p)/(4s)} \\
&\leq C (\log n/n)^{\alpha_1 p}.
\end{aligned}$$

• Si  $\epsilon < 0$ .

– *Majoration de  $T_1$ .* On procède de manière identique à la majoration du terme  $T_2$  pour le cas  $\epsilon > 0$ . Puisque  $\epsilon < 0$ , il vient

$$\begin{aligned}
T_1 &\leq C (\log n)^{(p-\pi)/2} n^{(\pi-p)/2} \sum_{m \in \mathbb{N}} 2^{-m\pi} \left( \sum_{j=\tau}^{j_4} 2^{j(1/2-1/p)} 2^{\delta j((p-\pi)/p)} 2^{-j(s+1/2-1/\pi)\pi/p} \right)^p \\
&\leq C (\log n)^{(p-\pi)/2} n^{(\pi-p)/2} \sum_{m \in \mathbb{N}} 2^{-m\pi} \left( \sum_{j=\tau}^{j_4} 2^{-j\epsilon/p} \right)^p \\
&\leq C (\log n)^{(p-\pi)/2} n^{(\pi-p)/2} \sum_{m \in \mathbb{N}} 2^{-m\pi} 2^{-j_4\epsilon} \\
&\leq C (\log n/n)^{\alpha_2 p} \sum_{m \in \mathbb{N}} 2^{-m\pi/2+m(2\delta+1)(\pi-p)/(4s)} \leq C (\log n/n)^{\alpha_2 p}.
\end{aligned}$$

– *Majoration de  $T_2$ .* En utilisant une propriété élémentaire de la base d'ondelettes en norme  $\mathbb{L}^p$  (2.6) et l'inclusion  $B_{\pi,r}^s(L) \subseteq B_{p,\infty}^{s-1/\pi+1/p}(L)$ , on a

$$\begin{aligned}
T_2 &\leq C \sum_{m \in \mathbb{N}} 2^{-mp} \left( \sum_{j=j_4+1}^{\infty} 2^{j(1/2-1/p)} \left( \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \right)^{1/p} \right)^p \\
&\leq C \sum_{m \in \mathbb{N}} 2^{-mp} 2^{-j_4(s-1/\pi+1/p)p} \leq C (\log n/n)^{\alpha_2 p} \sum_{m \in \mathbb{N}} 2^{-mp/2+(m/2s)(p/\pi-1)} \\
&\leq C (\log n/n)^{\alpha_2 p}.
\end{aligned}$$

On en déduit que

$$\mathcal{Q}_1^*(f) \leq C (\log n/n)^{\alpha_2 p}.$$

- Si  $\epsilon = 0$ .

La majoration du terme  $T_2$  obtenue précédemment est toujours valable. Il nous reste à analyser la majoration du terme  $T_1$ . En reprenant les notations précédentes et en utilisant l'inégalité (5.23), il vient

$$T_1 \leq C n^{(\pi-p)/2} (\log n)^{(p-\pi)/2} \sum_{m \in \mathbb{N}} 2^{-m\pi} \left( \sum_{j=\tau}^{j_4} \mathcal{L}_j \right)^p,$$

où

$$\mathcal{L}_j = (2^{j(s+1/2-1/\pi)\pi} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^\pi)^{1/p}.$$

Distinguons le cas où  $\pi \geq rp$  et le cas où  $\pi < rp$ .

- Si  $\pi \geq rp$ . L'inclusion  $B_{\pi,r}^s(L) \subseteq B_{\pi,\pi/p}^s(L)$  implique  $\sum_{j=\tau}^{\infty} \mathcal{L}_j \leq C$  et a fortiori

$$T_1 \leq C n^{(\pi-p)/2} (\log n)^{(p-\pi)/2} \leq C (\log n/n)^{\alpha_2 p}.$$

- Si  $\pi < rp$ . En utilisant l'inégalité de Hölder et l'inclusion  $f \in B_{\pi,r}^s(L) \subseteq B_{\pi,\infty}^s(L)$ , on a  $\mathcal{L}_j \leq L$  et  $(\sum_{j=\tau}^{\infty} \mathcal{L}_j^{pr/\pi})^{\pi/r} \leq L$ . Cela entraîne

$$\begin{aligned} \sum_{m \in \mathbb{N}} 2^{-m\pi} \left( \sum_{j=j_1}^{j_4} \mathcal{L}_j \right)^p &\leq \sum_{m \in \mathbb{N}} 2^{-m\pi} \left( \sum_{j=\tau}^{\infty} \mathcal{L}_j^{pr/\pi} \right)^{\pi/r} \left( \sum_{j=\tau}^{j_4} \mathcal{L}_j^{1/(1-\pi/(rp))} \right)^{p-\pi/r} \\ &\leq C \sum_{m \in \mathbb{N}} 2^{-m\pi} j_4^{(p-\pi/r)} \leq C (\log n)^{(p-\pi/r)}. \end{aligned}$$

D'où

$$T_1 \leq C (\log n)^{(p-\pi/r)} n^{(\pi-p)/2} (\log n)^{(p-\pi)/2} \leq C (\log n/n)^{\alpha_2 p} (\log n)^{(p-\pi/r)}.$$

On obtient les majorations désirées.

- *Majoration de  $\mathcal{Q}_2(f)$ .* L'inégalité de Minkowski, une propriété élémentaire de la base d'ondelettes en norme  $\mathbb{L}^p$  (2.6), l'inclusion  $B_{\pi,r}^s(L) \subseteq B_{p,r}^{s-1/\pi+1/p}(L)$  et le fait que  $s > 1/\pi - \delta - 1/2 + 1/(2\nu)$  entraînent

$$\begin{aligned} \mathcal{Q}_2(f) &\leq C \left( \sum_{j=j_2+1}^{\infty} 2^{j(1/2-1/p)} \left( \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \right)^{1/p} \right)^p \leq C \left( \sum_{j=j_2+1}^{\infty} 2^{-j(s-1/\pi+1/p)} \right)^p \\ &\leq C 2^{-j_2(s-1/\pi+1/p)p} \leq C (n^{-\alpha_1 p} \wedge (\log n/n)^{\alpha_2 p}). \end{aligned} \quad (5.24)$$

En combinant les majorations précédentes, on obtient les bornes supérieures souhaitées. La preuve du Théorème 5.4.2 est complète.  $\square$

**Preuve de la Proposition 5.4.2.** Supposons que les hypothèses (F1) et (F3) soient satisfaites. Pour  $f \in \mathbb{L}^p([0, 1])$ ,  $\mu = \mu_2$  et  $\alpha \in ]0, 1[$ , Un résultat de Kerkyacharian et Picard (2000, Theorem 5.2) nous assure l'existence d'une constante  $C > 0$  telle que

$$\mathcal{G}_1(f) \leq C(\mathbb{E}_f^n(\|\hat{f}_n^h - f\|_p^p) + n^{-\alpha p/2}), \quad (5.25)$$

où

$$\mathcal{G}_1(f) = \left\| \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} 1_{\{|\beta_{j,k}| \leq 2^{\delta j} 2^{-1} \mu_2 \sqrt{(\log n/n)}\}} \psi_{j,k} \right\|_p^p.$$

Le reste de la preuve est inspiré par une démonstration de Cai (2002b, Proof of Theorem 1). Soit  $j_3$  un entier tel que

$$c_0 (n/\log n)^{1/(2(s+\delta)+1)} \leq 2^{j_3} \leq C_0 (n/\log n)^{1/(2(s+\delta)+1)},$$

où  $c_0 > 0$  et  $C_0 > 0$  désignent deux constantes choisies a posteriori. Considérons la fonction  $f_* : [0, 1] \rightarrow \mathbb{R}$  définie par

$$f_*(x) = L_* 2^{-j_3(s+1/2)} \sum_{k=0}^{2^{j_3}-1} \psi_{j_3,k}(x), \quad (5.26)$$

où  $L_* > 0$  est une constante choisie a posteriori. Les coefficients d'ondelettes de  $f_*$  valent

$$\beta_{j,k} = \int_0^1 f_*(x) \psi_{j,k}(x) dx = \begin{cases} L_* 2^{-j_3(s+1/2)}, & \text{si } j = j_3, \\ 0, & \text{sinon.} \end{cases}$$

Par conséquent, on a  $2^{j_3(s+1/2)} (2^{-j_3} \sum_{k=0}^{2^{j_3}-1} |\beta_{j_3,k}|^\pi)^{1/\pi} = L_*$ . Avec un choix convenable de  $L_*$ , on a donc  $f_* \in B_{\pi,r}^s(L)$ .

En prenant  $c_0$  suffisamment grand, on a

$$|\beta_{j_3,k}| = L_* 2^{-j_3(s+1/2)} \leq L_* c_0^{-j_3(s+1/2+\delta)} 2^{\delta j_3} \sqrt{(\log n/n)} \leq 2^{-1} \mu_2 2^{\delta j_3} \sqrt{(\log n/n)}.$$

Une propriété élémentaire de la base d'ondelettes en norme  $\mathbb{L}^p$  (2.6) entraîne

$$\begin{aligned} \mathcal{G}_1(f_*) &= L_*^p 2^{-j_3(s+1/2)p} \left\| \sum_{k=0}^{2^{j_3}-1} \psi_{j_3,k} \right\|_p^p \geq c 2^{-j_3(s+1/2)p} 2^{j_3 p/2} = c 2^{-j_3 s p} \\ &\geq c C_0^{-s p} (\log n/n)^{\alpha_1 p}. \end{aligned} \quad (5.27)$$

En mettant les inégalités (5.25) et (5.27) ensemble, en prenant  $\alpha = 2\alpha_1$  et  $n$  grand, on montre l'existence d'une constante  $c > 0$  telle que

$$\sup_{f \in B_{\pi,r}^s(L)} (\log n/n)^{-\alpha_1 p} \mathbb{E}_f^n(\|\hat{f}_n^h - f\|_p^p) \geq (\log n/n)^{-\alpha_1 p} \mathbb{E}_f^n(\|\hat{f}_n^h - f_*\|_p^p) \geq c.$$

Cela termine la preuve de la Proposition 5.4.2. □

### 5.7.3 Preuves des résultats de la Section 5.5

**Preuve du Théorème 5.5.1.** Soit  $p \geq 2$  et  $\alpha \in ]0, 1[$ . En premier lieu, on montre l'inclusion  $\mathcal{M}(\hat{f}_n^g, p, n^{-\alpha p/2}) \subseteq \mathcal{W}_\delta^*((1-\alpha)p, p)$ . Puis, on montre l'inclusion  $\mathcal{W}_\delta^*((1-\alpha)p, p) \subseteq \mathcal{M}(\hat{f}_n^g, p, n^{-\alpha p/2})$ .

• *Inclusion  $\mathcal{M}(\hat{f}_n^g, p, n^{-\alpha p/2}) \subseteq \mathcal{W}_\delta^*((1-\alpha)p, p)$ .* Si  $f \in \mathcal{M}(\hat{f}_n^g, p, n^{-\alpha p/2})$  alors le Théorème 5.3.1 justifie l'existence d'une constante  $C > 0$  vérifiant

$$\mathcal{Q}_1 = \left\| \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} 1_{\{b_j \leq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} \psi_{j,k} \right\|_p^p \leq C n^{-\alpha p/2}. \quad (5.28)$$

Soit  $c$  une constante petite. Si  $u > c$  alors la propriété d'inconditionnalité (2.3) implique

$$\left\| \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} 1_{\{b_j \leq 2^{\delta j} u\}} \psi_{j,k} \right\|_p^p \leq C \|f\|_p^p \leq C u^{\alpha p}.$$

Si  $u \leq c$  alors il existe un entier  $n$  tel que  $2^{-1} \mu_1 (n+1)^{-1/2} < u \leq 2^{-1} \mu_1 n^{-1/2}$ . Par la propriété d'inconditionnalité (2.3) et l'inégalité (5.28), on a

$$\begin{aligned} \left\| \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} 1_{\{b_j \leq 2^{\delta j} u\}} \psi_{j,k} \right\|_p^p &\leq \left\| \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} 1_{\{b_j \leq 2^{\delta j} 2^{-1} \mu_1 n^{-1/2}\}} \psi_{j,k} \right\|_p^p \\ &\leq C n^{-\alpha p/2} \leq C u^{\alpha p} (n/n+1)^{-\alpha p/2} \leq C u^{\alpha p}. \end{aligned}$$

Par conséquent,  $f \in \mathcal{W}_\delta^*((1-\alpha)p, p)$ . D'où l'inclusion

$$\mathcal{M}(\hat{f}_n^g, p, n^{-\alpha p/2}) \subseteq \mathcal{W}_\delta^*((1-\alpha)p, p).$$

• *Inclusion  $\mathcal{W}_\delta^*((1-\alpha)p, p) \subseteq \mathcal{M}(\hat{f}_n^g, p, n^{-\alpha p/2})$ .* L'objectif est d'appliquer le Théorème 5.3.2. Soit  $j_2$  est un entier tel que

$$2^{j_2} \geq c n^\nu,$$

où  $c$  désigne une constante choisie a posteriori. Si  $f \in \mathcal{W}_\delta^*((1-\alpha)p, p)$  avec  $\alpha \in ]1 - (2\delta + 1)\nu, 1[$  alors, pour des constantes  $c$  et  $c_*$  convenablement choisies, on a

$$\begin{aligned} n^{\nu p(\delta+1/2)} \sum_{j=j_2+1}^{\infty} 1_{\{b_j \geq c_* n^{-1/2} 2^{\delta j}\}} &\leq c^{-1} \sum_{j=j_2+1}^{\infty} 1_{\{b_j \geq c_* n^{-1/2} 2^{\delta j}\}} 2^{jp(\delta+1/2)} \|\psi\|_p^p \\ &\leq c^{-1} \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} 1_{\{b_j \geq c_* n^{-1/2} 2^{\delta j}\}} 2^{\delta j p} \|\psi_{j,k}\|_p^p \\ &\leq c^{-1} C n^{p/2} \sum_{m \in \mathbb{N}} 2^{-mp} \left\| \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} 1_{\{b_j \leq c_* n^{-1/2} 2^{\delta j} 2^{m+1}\}} \psi_{j,k} \right\|_p^p \\ &\leq n^{(1-\alpha)p/2} c^{-1} C \sum_{m \in \mathbb{N}} 2^{(\alpha-1)mp} < n^{\nu p(\delta+1/2)}. \end{aligned}$$

## 5.7. DÉMONSTRATIONS

Le passage de la deuxième à la troisième ligne s'est faite grâce à une inégalité similaire à (5.17). Par conséquent, pour tout  $j \in \{j_2 + 1, \dots, \infty\}$ , on a

$$b_j \leq c_* n^{-1/2} 2^{\delta j}.$$

Si  $f \in \mathcal{W}_\delta^*((1-\alpha)p, p)$  alors la propriété d'inconditionnalité (2.3) entraîne

$$\begin{aligned} \mathcal{Q}_2(f) &= \left\| \sum_{j=j_2+1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k} \right\|_p^p \leq \left\| \sum_{j=j_2+1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} 1_{\{b_j \leq c_* n^{-1/2} 2^{\delta j}\}} \psi_{j,k} \right\|_p^p \\ &\leq \left\| \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} 1_{\{b_j \leq c_* n^{-1/2} 2^{\delta j}\}} \psi_{j,k} \right\|_p^p \leq C n^{-\alpha p/2}. \end{aligned}$$

De plus, si  $f \in \mathcal{W}_\delta^*((1-\alpha)p, p)$  alors

$$\begin{aligned} \mathcal{Q}_1^*(f) &= \sum_{m \in \mathbb{N}} 2^{-mp} \left\| \sum_{j=j_1}^{j_2} \sum_{k=0}^{2^j-1} \beta_{j,k} 1_{\{b_j \leq 2^{-1} \mu_1 n^{-1/2} 2^{\delta j} 2^{m+1}\}} \psi_{j,k} \right\|_p^p \\ &\leq C n^{-\alpha p/2} \sum_{m \in \mathbb{N}} 2^{(\alpha-1)mp} \leq C n^{-\alpha p/2}. \end{aligned}$$

En vertu du Théorème 5.3.2, pour  $\alpha \in ]1 - (2\delta + 1)\nu, 1[$ , il existe une constante  $C > 0$  telle que

$$\mathbb{E}_f^n(\|\hat{f}_n^g - f\|_p^p) \leq C n^{-\alpha p/2}.$$

On en déduit que  $f \in \mathcal{M}(\hat{f}_n^g, p, n^{-\alpha p/2})$ . D'où l'inclusion

$$\mathcal{W}_\delta^*((1-\alpha)p, p) \subseteq \mathcal{M}(\hat{f}_n^g, p, n^{-\alpha p/2}).$$

Cela termine la preuve du Théorème 5.5.1. □

**Preuve du Théorème 5.5.2.** Soit  $p \geq 2$  et  $\alpha \in ]0, 1[$ . Si  $f \in \mathcal{M}(\hat{f}_n^h, p, n^{-\alpha p/2})$  alors un résultat de Kerkycharian et Picard (2000, Theorem 5.2) implique

$$\mathcal{G}_1(f) = \left\| \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} 1_{\{|\beta_{j,k}| \leq 2^{\delta j} 2^{-1} \mu_2 \sqrt{\log n/n}\}} \psi_{j,k} \right\|_p^p \leq C \mathbb{E}_f^n(\|\hat{f}_n^h - f\|_p^p) \leq C n^{-\alpha p/2}$$

et

$$\mathcal{G}_2(f) = \left\| \sum_{j=j_2+1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k} \right\|_p^p \leq \mathbb{E}_f^n(\|\hat{f}_n^h - f\|_p^p) \leq C n^{-\alpha p/2}.$$

Pour conclure, il suffit de montrer que ces inégalités entraînent que les termes  $\mathcal{Q}_1(f)$  et  $\mathcal{Q}_2(f)$  du Théorème 5.3.2 sont majorés par  $C n^{-\alpha p/2}$ .

• *Majoration de  $\mathcal{Q}_2(f)$ .* On a

$$\mathcal{Q}_2(f) = \mathcal{G}_2(f) \leq C n^{-\alpha p/2}. \tag{5.29}$$



• *Majoration de  $\mathcal{Q}_1^*(f)$ .*

– Montrons que si  $\mathcal{G}_1(f) \leq Cn^{-\alpha p/2}$  alors on a

$$\sup_{u>0} \eta(u)^{r-p} \left\| \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} 1_{\{|\beta_{j,k}| \leq 2^{\delta j} u\}} \psi_{j,k} \right\|_p^p < \infty, \quad (5.30)$$

où  $\eta : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  est la fonction continue croissante définie par

$$\eta(u) = \begin{cases} u \log((u \wedge \exp(-p/2))^{-1})^{-1/2}, & u > 0, \\ 0, & u = 0. \end{cases}$$

Soit  $c$  une constante petite. Si  $u > c$  alors la propriété d'inconditionnalité (2.3) donne

$$\left\| \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} 1_{\{|\beta_{j,k}| \leq 2^{\delta j} u\}} \psi_{j,k} \right\|_p^p \leq C \|f\|_p^p \leq C \eta(u)^{\alpha p}.$$

Si  $u \leq c$  alors il existe un entier  $n$  tel que

$$2^{-1} \mu_2 \sqrt{(\log(n+1)/(n+1))} < u \leq 2^{-1} \mu_2 \sqrt{(\log n/n)}.$$

Par conséquent, si  $\mathcal{G}_1(f) \leq Cn^{-\alpha p/2}$  alors on a

$$\begin{aligned} \left\| \sum_{j=j_1}^{j_2} \sum_{k=0}^{2^j-1} \beta_{j,k} 1_{\{|\beta_{j,k}| \leq 2^{\delta j} u\}} \psi_{j,k} \right\|_p^p &\leq \left\| \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} 1_{\{|\beta_{j,k}| \leq 2^{\delta j} 2^{-1} \mu_2 \sqrt{(\log n/n)}\}} \psi_{j,k} \right\|_p^p \\ &\leq C \eta(2^{-1} \mu_2 \sqrt{(\log n/n)})^{\alpha p} \\ &\leq C \eta(2^{-1} \mu_2 \sqrt{(\log n/n)})^{\alpha p} \dots \\ &\dots \eta(2^{-1} \mu_2 \sqrt{(\log(n+1)/n+1)})^{-\alpha p} \eta(u)^{\alpha p} \\ &\leq C \eta(u)^{\alpha p}. \end{aligned}$$

L'inégalité (5.30) est prouvée.

– Grâce à la taille des blocs, pour tout  $k$  dans  $\mathcal{B}_{j,K}$ , il existe une constante  $C > 0$  vérifiant

$$\{b_{j,K} \leq \mu_1 2^{m+1} n^{-1/2} 2^{\delta j}\} \subseteq \{|\beta_{j,k}| \leq C \mu_1 2^{m+1} 2^{\delta j} \sqrt{(\log n/n)}\}.$$

Si  $\mathcal{G}_1(f) \leq Cn^{-\alpha p/2}$  alors la propriété d'inconditionnalité (2.3) et un résultat prouvé

par Kerkyacharian et al. (2005, Lemma 2) nous donnent

$$\begin{aligned}
 \mathcal{Q}_1^*(f) &= \sum_{m \in \mathbb{N}} 2^{-mp} \left\| \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\{b_{j,K} \leq \mu_1 2^{m+1} n^{-1/2} 2^{\delta j}\}} \psi_{j,k} \right\|_p^p \\
 &\leq \sum_{m \in \mathbb{N}} 2^{-mp} \left\| \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} 1_{\{|\beta_{j,k}| \leq C \mu_1 2^{m+1} 2^{\delta j} \sqrt{(\log n/n)}\}} \psi_{j,k} \right\|_p^p \\
 &\leq C \sum_{m \in \mathbb{N}} 2^{-mp} \eta(\sqrt{(\log n/n)} C \mu_1 2^{m+1})^{\alpha p} \\
 &\leq C (\log n/n)^{\alpha p/2} (\log((C \mu_1 \sqrt{(\log n/n)} \wedge v)^{-1}))^{-\alpha p/2} \\
 &\leq C n^{-\alpha p/2}.
 \end{aligned} \tag{5.31}$$

Le Théorème 5.3.2 et les inégalités (5.29) et (5.31) justifient l'existence d'une constante  $C > 0$  telle que

$$\mathbb{E}_f^n(\|\hat{f}_n^b - f\|) \leq C(n^{-\alpha p/2} + \mathcal{Q}_1^*(f) + \mathcal{Q}_2(f)) \leq C n^{-\alpha p/2}.$$

On en déduit que  $f \in \mathcal{M}(\hat{f}_n^b, p, n^{-\alpha p/2})$ . D'où l'inclusion

$$\mathcal{M}(\hat{f}_n^h, p, n^{-\alpha p/2}) \subseteq \mathcal{M}(\hat{f}_n^b, p, n^{-\alpha p/2}).$$

Maintenant, montrons l'inclusion stricte. Soit  $j_4$  un entier tel que

$$c_0 (n/\log n)^{(1-\alpha)/(1+2\delta)} \leq 2^{j_4} \leq C_0 (n/\log n)^{(1-\alpha)/(1+2\delta)},$$

où  $c_0 > 0$  et  $C_0 > 0$  désignent deux constantes choisies a posteriori. L'hypothèse  $\alpha \in ]1 - \nu(1 + 2\delta), 1[$  nous assure que  $j_4 \leq j_2$ . Considérons la fonction  $f_* : [0, 1] \rightarrow \mathbb{R}$  définie par

$$f_*(x) = L_* 2^{-j_4(2\delta\alpha+1)/(2(1-\alpha))} \sum_{k=0}^{2^{j_4}-1} \psi_{j_4,k}(x),$$

où  $L_* > 0$  est une constante choisie a posteriori. En procédant de la même manière que pour la Proposition 5.4.2, on montre que  $f_* \in B_{\pi,r}^{\alpha(1+2\delta)/(2(1-\alpha))}(L)$  pour tout  $\alpha \in ]0, 1[, \pi$  et  $r$ .

En particulier, on a  $f_* \in B_{p,r}^{\alpha(1+2\delta)/(2(1-\alpha))}(L)$ . Par la Proposition 5.4.1, il vient  $f_* \in \mathcal{M}(\hat{f}_n^b, p, n^{-\alpha p/2})$ .

Or, par des calculs similaires à la Proposition 5.4.2, on montre que l'existence d'une constante  $c > 0$  telle que

$$\mathbb{E}_f^n(\|\hat{f}_n^h - f_*\|_p^p) \geq c(\log n)^{\alpha p/2} n^{-\alpha p/2}.$$

Par conséquent,  $f_* \notin \mathcal{M}(\hat{f}_n^h, p, n^{-\alpha p/2})$ . On en déduit l'inclusion stricte

$$\mathcal{M}(\hat{f}_n^h, p, n^{-\alpha p/2}) \subset \mathcal{M}(\hat{f}_n^b, p, n^{-\alpha p/2}).$$

Cela termine la preuve du Théorème 5.5.2. □

### 5.7.4 Preuves des résultats de la Section 5.6

Les preuves qui suivent reposent sur deux lemmes auxiliaires : l'inégalité de Talagrand et l'inégalité de Cirelson.

LEMME 5.7.1 (TALAGRAND (1994))

Soient  $(V_1, \dots, V_n)$  des variables aléatoires i.i.d et  $(\epsilon_1, \dots, \epsilon_n)$  des variables aléatoires de Rademacher i.i.d indépendantes des  $(V_1, \dots, V_n)$ . Soit  $\mathcal{F}$  une classe de fonctions uniformément bornées par  $T$ . Définissons l'opérateur  $r_n : \mathcal{F} \rightarrow \mathbb{R}$  par

$$r_n(h) = n^{-1} \sum_{i=1}^n h(V_i) - \mathbb{E}(h(V_1)).$$

Supposons que

$$\sup_{h \in \mathcal{F}} \text{Var}(h(V_1)) \leq v \quad \text{et} \quad \mathbb{E}(\sup_{h \in \mathcal{F}} \sum_{i=1}^n \epsilon_i h(V_i)) \leq nH.$$

Alors il existe deux constantes  $C_1 > 0$  et  $C_2 > 0$  telles que, pour tout  $t > 0$ , on a

$$\mathbb{P}(\sup_{h \in \mathcal{F}} r_n(h) \geq t + C_2 H) \leq \exp(-nC_1 (t^2 v^{-1} \wedge tT^{-1})).$$

LEMME 5.7.2 (CIRELSON ET AL. (1976))

Soient  $\mathcal{D} \subseteq \mathbb{R}$  et  $(\eta_t)_{t \in \mathcal{D}}$  un processus gaussien centré. Supposons que

$$\mathbb{E}(\sup_{t \in \mathcal{D}} \eta_t) \leq N \quad \text{et} \quad \sup_{t \in \mathcal{D}} \text{Var}(\eta_t) \leq Q.$$

Alors, pour tout  $x > 0$ , on a

$$\mathbb{P}(\sup_{t \in \mathcal{D}} \eta_t \geq x + N) \leq \exp(-x^2/(2Q)). \quad (5.32)$$

**Preuve du Théorème 5.6.2.** Pour la preuve des hypothèses (F1) et (F3), voir la preuve du Théorème 3.4.1. Il reste à étudier l'hypothèse (F2).

Pour  $1 \leq p < 2$ , une inégalité de norme  $l_p$  entraîne

$$\mathbb{P}_f^n((l_j^{-1} \sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^p)^{1/p} \geq 2^{-1} \mu n^{-1/2}) \leq \mathbb{P}_f^n((l_j^{-1} \sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^2)^{1/2} \geq 2^{-1} \mu n^{-1/2}).$$

Par conséquent, il suffit de montrer l'hypothèse (F2) pour  $p \geq 2$ .

Par définition de  $\hat{\beta}_{j,k}$ , on a la décomposition  $\hat{\beta}_{j,k} - \beta_{j,k} = A_{j,k} + B_{j,k}$  où

$$A_{j,k} = n^{-1} \sum_{i=1}^n f(X_i) g(X_i)^{-1} \psi_{j,k}(X_1) - \mathbb{E}_f^n(f(X_1) g(X_1)^{-1} \psi_{j,k}(X_1))$$

et

$$B_{j,k} = n^{-1} \sum_{i=1}^n g(X_i)^{-1} \psi_{j,k}(X_i) z_i.$$

En vertu de l'inégalité de Minkowski  $l_p$ , pour tout  $\mu > 0$ , on a

$$\mathbb{P}_f^n((l_j^{-1} \sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^p)^{1/p} \geq 2^{-1} \mu n^{-1/2}) \leq \mathcal{U} + \mathcal{V},$$

où

$$\mathcal{U} = \mathbb{P}_f^n((l_j^{-1} \sum_{(K)} |A_{j,k}|^p)^{1/p} \geq 4^{-1} \mu n^{-1/2})$$

et

$$\mathcal{V} = \mathbb{P}_f^n((l_j^{-1} \sum_{(K)} |B_{j,k}|^p)^{1/p} \geq 4^{-1} \mu n^{-1/2}).$$

Étudions les majorations de  $\mathcal{U}$  et  $\mathcal{V}$ .

• *Majoration de  $\mathcal{U}$ .* Ici, l'objectif est d'appliquer l'inégalité de Talagrand décrite dans le Lemme 5.7.1. Pour ce faire, considérons l'ensemble  $\mathcal{C}_q$  défini par

$$\mathcal{C}_q = \{a = (a_{j,k}) \in \mathbb{Z}^*; \quad \sum_{(K)} |a_{j,k}|^q \leq 1\}, \quad (5.33)$$

et la classe de fonctions  $\mathcal{F}$  définie par

$$\mathcal{F} = \{h; \quad h(x) = f(x)g(x)^{-1} \sum_{(K)} a_{j,k} \psi_{j,k}(x), \quad a \in \mathcal{C}_q\}.$$

Par un argument de dualité, on a

$$(\sum_{(K)} |A_{j,k}|^p)^{1/p} = \sup_{a \in \mathcal{C}_q} \sum_{(K)} a_{j,k} A_{j,k} = \sup_{h \in \mathcal{F}} r_n(h),$$

où  $r_n$  désigne l'opérateur défini dans le Lemme 5.7.1.

Il reste à identifier les paramètres  $T$ ,  $H$  et  $v$  inhérents à l'inégalité de Talagrand.

Avant toute chose, remarquons que pour  $p \geq 2$ , on a  $q = 1 + (p-1)^{-1} \leq 2$ . Une inégalité de norme  $l_p$  implique

$$\sup_{a \in \mathcal{C}_q} (\sum_{(K)} |a_{j,k}|^2)^{1/2} \leq \sup_{a \in \mathcal{C}_q} (\sum_{(K)} |a_{j,k}|^q)^{1/q} \leq 1.$$

– *Evaluation de  $T$ .* Soit  $h \in \mathcal{F}$ . Par l'inégalité de Hölder, les hypothèses de bornitude faites sur  $f$  et  $g$  et la propriété de concentration (2.5), il vient

$$\begin{aligned} |h(x)| &\leq |f(x)| |g(x)|^{-1} (\sum_{(K)} |\psi_{j,k}(x)|^2)^{1/2} \sup_{a \in \mathcal{C}_q} (\sum_{(K)} |a_{j,k}|^2)^{1/2} \\ &\leq \|f\|_\infty \|1/g\|_\infty (\sum_{(K)} |\psi_{j,k}(x)|^2)^{1/2} \leq C 2^{j/2}, \quad x \in [0, 1]. \end{aligned}$$

D'où  $T = C2^{j/2}$ .

– *Evaluation de  $H$ .* L'inégalité de Hölder  $l_p$  et l'inégalité de Hölder impliquent

$$\begin{aligned}
 & \mathbb{E}_f^n \left( \sup_{a \in \mathcal{C}_q} \sum_{i=1}^n \left( \sum_{(K)} a_{j,k} \epsilon_i f(X_i) g(X_i)^{-1} \psi_{j,k}(X_i) \right) \right) \\
 & \leq \sup_{a \in \mathcal{C}_q} \left( \sum_{(K)} |a_{j,k}|^q \right)^{1/q} \left( \sum_{(K)} \mathbb{E}_f^n \left( \left| \sum_{i=1}^n \epsilon_i f(X_i) g(X_i)^{-1} \psi_{j,k}(X_i) \right|^p \right) \right)^{1/p} \\
 & \leq \left( \sum_{(K)} \mathbb{E}_f^n \left( \left| \sum_{i=1}^n \epsilon_i f(X_i) g(X_i)^{-1} \psi_{j,k}(X_i) \right|^p \right) \right)^{1/p}.
 \end{aligned} \tag{5.34}$$

Puisque  $(\epsilon_1, \dots, \epsilon_n)$  sont des variables de Rademacher i.i.d indépendantes des variables  $(X_1, \dots, X_n) = \mathbb{X}$ , l'inégalité de Khintchine donne

$$\begin{aligned}
 \mathbb{E}_f^n \left( \left| \sum_{i=1}^n \epsilon_i f(X_i) g(X_i)^{-1} \psi_{j,k}(X_i) \right|^p \right) &= \mathbb{E}_f^n \left( \mathbb{E}_f^n \left( \left| \sum_{i=1}^n \epsilon_i f(X_i) g(X_i)^{-1} \psi_{j,k}(X_i) \right|^p \middle| \mathbb{X} \right) \right) \\
 &\leq C \mathbb{E}_f^n \left( \left| \sum_{i=1}^n |f(X_i)|^2 |g(X_i)|^{-2} |\psi_{j,k}(X_i)|^2 \right|^{p/2} \right) \\
 &= CI.
 \end{aligned} \tag{5.35}$$

Considérons les variables aléatoires i.i.d  $(N_1, \dots, N_n)$  avec

$$N_i = |f(X_i)|^2 |g(X_i)|^{-2} |\psi_{j,k}(X_i)|^2, \quad i \in \{1, \dots, n\}.$$

Une inégalité élémentaire de convexité entraîne  $I \leq 2^{p/2-1}(I_1 + I_2)$  où

$$I_1 = \mathbb{E}_f^n \left( \left| \sum_{i=1}^n (N_i - \mathbb{E}_f^n(N_1)) \right|^{p/2} \right) \quad \text{et} \quad I_2 = n^{p/2} \mathbb{E}_f^n(N_1)^{p/2}.$$

Étudions les majorations des termes  $I_1$  et  $I_2$ .

– *Majoration de  $I_1$ .* L'inégalité de Rosenthal appliquée aux variables  $(N_1, \dots, N_n)$  et l'inégalité de Cauchy-Schwartz entraînent

$$\begin{aligned}
 I_1 &\leq C(n \mathbb{E}_f^n(|N_1 - \mathbb{E}_f^n(N_1)|^{p/2}) + (n \mathbb{E}_f^n(|N_1 - \mathbb{E}_f^n(N_1)|^2))^{p/4}) \\
 &\leq C(n \mathbb{E}_f^n(|N_1|^{p/2}) + (n \mathbb{E}_f^n(|N_1|^2))^{p/4}).
 \end{aligned}$$

Pour tout  $m \geq 1$ ,  $j \in \{j_1, \dots, j_2\}$  et  $k \in \{0, \dots, 2^j - 1\}$ , les hypothèses de bornitude faites sur  $f$  et  $g$  entraînent

$$\begin{aligned}
 \mathbb{E}_f^n(|N_1|^m) &= \int_0^1 |f(x)|^{2m} |g(x)|^{2m-1} |\psi_{j,k}(x)|^{2m} dx \\
 &\leq \|f\|_\infty^{2m} \|1/g\|_\infty^{-2m+1} \|\psi\|_\infty^{2m-2} 2^{j(m-1)} \int_0^1 |\psi_{j,k}(x)|^2 dx \leq C 2^{j_2(m-1)} \\
 &\leq C n^{m-1}.
 \end{aligned}$$

## 5.7. DÉMONSTRATIONS

---

Il s'ensuit  $I_1 \leq Cn^{p/2}$ .

– *Majoration de  $I_2$ .* Comme  $\mathbb{E}_f^n(N_1) \leq C$ , on a  $I_2 \leq Cn^{p/2}$ .

Les majorations obtenues impliquent

$$I \leq C(I_1 + I_2) \leq Cn^{p/2}. \quad (5.36)$$

En combinant les inégalités (5.34), (5.35) et (5.36), il vient

$$\mathbb{E}_f^n \left( \sup_{a \in \mathcal{C}_q} \sum_{i=1}^n \left( \sum_{(K)} a_{j,k} \epsilon_i f(X_i) g(X_i)^{-1} \psi_{j,k}(X_i) \right) \right) \leq \left( \sum_{(K)} I \right)^{1/p} \leq Cn^{1/2} l_j^{1/p}.$$

D'où  $H = Cn^{-1/2} l_j^{1/p}$ .

– *Evaluation de  $v$ .* En utilisant les hypothèses de bornitude faites sur  $f$  et  $g$  et l'orthonormalité de  $\zeta$ , on obtient

$$\begin{aligned} & \sup_{h \in \mathcal{F}} \text{Var}(h(X_1)) \\ & \leq \sup_{a \in \mathcal{C}_q} \mathbb{E}_f^n (|f(X_1)|^2 |g(X_1)|^{-2} \left| \sum_{(K)} a_{j,k} \psi_{j,k}(X_1) \right|^2) \\ & \leq \|f\|_\infty^2 \|1/g\|_\infty \sup_{a \in \mathcal{C}_q} \mathbb{E}_f^n \left( \sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} a_{j,k} a_{j,k'} g(X_1)^{-1} \psi_{j,k}(X_1) \psi_{j,k'}(X_1) \right) \\ & = C \sup_{a \in \mathcal{C}_q} \sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} a_{j,k} a_{j,k'} \int_0^1 \psi_{j,k}(x) \psi_{j,k'}(x) dx \\ & = C \sup_{a \in \mathcal{C}_q} \sum_{(K)} |a_{j,k}|^2 \leq C. \end{aligned}$$

D'où  $v = C$ .

Pour tout  $j \in \{j_1, \dots, j_2\}$ , on a  $n2^j \leq n2^{j_2} \leq Cn^{3/2}(\log n)^{-1/2}$  et  $l_j^{1/p} \geq l_{j_1}^{1/p} \asymp (\log n)^{1/2}$ . Par conséquent, si  $t = 8^{-1} \mu l_j^{1/p} n^{-1/2}$  alors

$$(t^2 v^{-1} \wedge tT^{-1}) \geq C \left( \mu^2 (\log n/n) \wedge \mu \sqrt{(\log n/(n2^j))} \right) \geq C \mu^2 (\log n/n).$$

En considérant un tel  $t$  et en prenant  $\mu$  suffisamment grand, l'inégalité de Talagrand implique

$$\begin{aligned} \mathcal{U} &= \mathbb{P}_f^n \left( (l_j^{-1} \sum_{(K)} |A_{j,k}|^p)^{1/p} \geq 4^{-1} \mu n^{-1/2} \right) \\ &\leq \mathbb{P}_f^n \left( (l_j^{-1} \sum_{(K)} |A_{j,k}|^p)^{1/p} \geq 8^{-1} \mu n^{-1/2} + Cn^{-1/2} \right) \leq \mathbb{P}_f^n \left( \sup_{h \in \mathcal{F}} r_n(h) \geq t + C_2 H \right) \\ &\leq \exp(-nC_1 (t^2 v^{-1} \wedge tT^{-1})) \leq \exp(-nC \mu^2 (\log n/n)) \leq n^{-p}. \end{aligned}$$

On obtient la majoration souhaitée pour le terme  $\mathcal{U}$ .

• *Majoration de  $\mathcal{V}$ .* L'objectif est d'appliquer l'inégalité de Cirelson décrite dans le Lemme 5.7.2.

Considérons l'ensemble  $C_q$  défini en (5.33) et le processus  $\mathcal{Z}(a)$  par

$$\mathcal{Z}(a) = \sum_{(K)} a_{j,k} B_{j,k}.$$

Conditionnellement à  $\mathbb{X} = (X_1, \dots, X_n)$ ,  $\mathcal{Z}(a)$  est gaussien et centré. Par un argument de dualité, on a

$$\sup_{a \in C_q} \mathcal{Z}(a) = \sup_{a \in C_q} \sum_{(K)} a_{j,k} B_{j,k} = \left( \sum_{(K)} |B_{j,k}|^p \right)^{1/p}.$$

Il nous reste à étudier les majorations de  $\mathbb{E}_f^n(\sup_{a \in C_q} \mathcal{Z}(a) | \mathbb{X})$  et  $\sup_{a \in C_q} \text{Var}_f^n(\mathcal{Z}(a) | \mathbb{X})$ .

– *Majoration de  $\mathbb{E}_f^n(\sup_{a \in C_q} \mathcal{Z}(a) | \mathbb{X})$ .* Considérons l'ensemble  $\mathcal{B}_\mu$  défini par

$$\mathcal{B}_\mu = \left\{ \left| n^{-1} \sum_{i=1}^n g(X_i)^{-1} |\psi_{j,k}(X_i)|^2 - 1 \right| \geq \mu \right\}.$$

Travaillons sur l'ensemble  $\mathcal{B}_\mu^c$ . Par l'inégalité de Jensen version conditionnelle, le fait que  $\mathcal{Z}(a) | \mathbb{X} \sim \mathcal{N}(0, n^{-2} \sum_{i=1}^n |g(X_i)|^{-2} |\psi_{j,k}(X_i)|^2)$  et l'hypothèse de minoration faite sur  $g$ , il vient

$$\begin{aligned} \mathbb{E}_f^n(\sup_{a \in C_q} \mathcal{Z}(a) | \mathbb{X}) &\leq \left( \sum_{(K)} \mathbb{E}_f^n(|B_{j,k}|^p | \mathbb{X}) \right)^{1/p} \\ &= C \left( \sum_{(K)} (n^{-2} \sum_{i=1}^n |g(X_i)|^{-2} |\psi_{j,k}(X_i)|^2)^{p/2} \right)^{1/p} \\ &\leq C \|1/g\|_\infty \left( \sum_{(K)} (n^{-2} \sum_{i=1}^n g(X_i)^{-1} |\psi_{j,k}(X_i)|^2)^{p/2} \right)^{1/p} \\ &\leq C n^{-1/2} \left( \sum_{(K)} (n^{-1} \sum_{i=1}^n g(X_i)^{-1} |\psi_{j,k}(X_i)|^2 - 1 + 1)^{p/2} \right)^{1/p} \\ &\leq C n^{-1/2} \left( \sum_{(K)} (\mu + 1)^{p/2} \right)^{1/p} \leq C(\mu + 1)^{1/2} l_j^{1/p} n^{-1/2}. \end{aligned}$$

D'où  $N = N(\mathbb{X}) = C(\mu + 1)^{1/2} l_j^{1/p} n^{-1/2}$ .

– *Majoration de  $\sup_{a \in C_q} \text{Var}_f^n(\mathcal{Z}(a) | \mathbb{X})$ .* Définissons l'ensemble  $\mathcal{A}_\mu$  par

$$\mathcal{A}_\mu = \left\{ \sup_{a \in C_q} \left( \sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} a_{j,k} a_{j,k'} (n^{-1} \sum_{i=1}^n g(X_i)^{-1} \psi_{j,k}(X_i) \psi_{j,k'}(X_i)) - \sum_{(K)} |a_{j,k}|^2 \right) \geq \mu \right\}.$$

## 5.7. DÉMONSTRATIONS

---

Travaillons sur l'ensemble  $\mathcal{A}_\mu^c$ . En utilisant les hypothèses de bornitude faites sur  $g$ , on a

$$\begin{aligned} G &= \sup_{a \in \mathcal{C}_q} \left( \sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} a_{j,k} a_{j,k'} (n^{-1} \sum_{i=1}^n |g(X_i)|^{-2} \psi_{j,k}(X_i) \psi_{j,k'}(X_i)) \right) \\ &\leq C \left[ \sup_{a \in \mathcal{C}_q} \left( \sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} a_{j,k} a_{j,k'} (n^{-1} \sum_{i=1}^n g(X_i)^{-1} \psi_{j,k}(X_i) \psi_{j,k'}(X_i)) \right) - \dots \right. \\ &\quad \left. \sum_{(K)} |a_{j,k}|^2 \right) + \sup_{a \in \mathcal{C}_q} \sum_{(K)} |a_{j,k}|^2 \Big] \leq C(\mu + 1). \end{aligned}$$

Comme  $\mathbb{E}_f^n(z_i z_{i'}) = 1$  si  $i = i'$  et 0 sinon, il s'ensuit

$$\begin{aligned} \sup_{a \in \mathcal{C}_q} \text{Var}_f^n(\mathcal{Z}(a) | \mathbb{X}) &= \sup_{a \in \mathcal{C}_q} \mathbb{E}_f^n(|\mathcal{Z}(a)|^2 | \mathbb{X}) \\ &= \sup_{a \in \mathcal{C}_q} \mathbb{E}_f^n \left( \sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} a_{j,k} a_{j,k'} B_{j,k} B_{j,k'} | \mathbb{X} \right) \\ &= \sup_{a \in \mathcal{C}_q} (n^{-2} \sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} a_{j,k} a_{j,k'} \sum_{i=1}^n \sum_{i'=1}^n |g(X_i)|^{-2} \psi_{j,k}(X_i) \psi_{j,k'}(X_{i'}) \mathbb{E}_f^n(z_i z_{i'})) \\ &= n^{-1} \sup_{a \in \mathcal{C}_q} \left( \sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} a_{j,k} a_{j,k'} (n^{-1} \sum_{i=1}^n |g(X_i)|^{-2} \psi_{j,k}(X_i) \psi_{j,k'}(X_i)) \right) \\ &= n^{-1} G \leq C n^{-1} (\mu + 1). \end{aligned}$$

D'où  $Q = Q(\mathbb{X}) = C n^{-1} (\mu + 1)$ .

Les valeurs de  $N$  et  $Q$  obtenues vont nous permettre de conclure. En effet, pour tout  $x > 0$ , on a

$$\begin{aligned} &\mathbb{P}_f^n(\sup_{a \in \mathcal{C}_q} \mathcal{Z}(a) \geq x + C(1 + \mu)^{1/2} l_j^{1/p} n^{-1/2}) \\ &= \mathbb{E}_f^n(\mathbb{P}_f^n(\sup_{a \in \mathcal{C}_q} \mathcal{Z}(a) \geq x + C(1 + \mu)^{1/2} l_j^{1/p} n^{-1/2} | \mathbb{X})) \\ &\leq \mathbb{E}_f^n(\mathbb{P}_f^n(\sup_{a \in \mathcal{C}_q} \mathcal{Z}(a) \geq x + C(1 + \mu)^{1/2} l_j^{1/p} n^{-1/2} | \mathbb{X})(1_{\mathcal{B}_\mu} + 1_{\mathcal{B}_\mu^c})) \\ &\leq \mathbb{P}_f^n(\mathcal{B}_\mu) + \mathbb{E}_f^n(\mathbb{P}_f^n(\sup_{a \in \mathcal{C}_q} \mathcal{Z}(a) \geq x + N(\mathbb{X}) | \mathbb{X})). \end{aligned} \tag{5.37}$$

L'inégalité de Cirelson décrite au Lemme 5.7.2 implique

$$\mathbb{E}_f^n(\mathbb{P}_f^n(\sup_{a \in \mathcal{C}_q} \mathcal{Z}(a) \geq x + N(\mathbb{X}) | \mathbb{X})) \leq \mathbb{E}_f^n(\exp(-(x^2/(2Q(\mathbb{X}))))). \tag{5.38}$$

De plus, par définition de  $\mathcal{A}_\mu$ , on a

$$\begin{aligned} \mathbb{E}_f^n(\exp(-(x^2/(2Q(\mathbb{X})))) &= \mathbb{E}_f^n(\exp(-(x^2/(2Q(\mathbb{X}))))(1_{\mathcal{A}_\mu} + 1_{\mathcal{A}_\mu^c})) \\ &\leq \mathbb{P}_f^n(\mathcal{A}_\mu) + \exp(-nx^2/(2(\mu + 1))). \end{aligned} \tag{5.39}$$



En mettant les inégalités (5.37), (5.38) et (5.39) ensemble, en prenant  $x = 8^{-1}\mu l_j^{1/p}n^{-1/2}$  et  $\mu$  suffisamment grand, on a

$$\begin{aligned} \mathcal{V} &= \mathbb{P}_f^n(\sup_{a \in \mathcal{C}_q} \mathcal{Z}(a) \geq 4^{-1}\mu l_j^{1/p}n^{-1/2}) \\ &\leq \mathbb{P}_f^n(\sup_{a \in \mathcal{C}_q} \mathcal{Z}(a) \geq 8^{-1}\mu l_j^{1/p}n^{-1/2} + C(1+\mu)^{1/2}l_j^{1/p}n^{-1/2}) \\ &\leq C(\mathbb{P}_f^n(\mathcal{A}_\mu) + \mathbb{P}_f^n(\mathcal{B}_\mu) + \exp(-C\mu^2 l_j^{2/p}/(\mu+1))). \end{aligned} \quad (5.40)$$

Le Lemme 5.7.3 ci-dessous étudie les majorations de  $\mathbb{P}_f^n(\mathcal{A}_\mu)$  et  $\mathbb{P}_f^n(\mathcal{B}_\mu)$ .

#### LEMME 5.7.3

Pour  $\mu$  suffisamment grand, il existe une constante  $C > 0$  telle que

$$\mathbb{P}_f^n(\mathcal{A}_\mu) \vee \mathbb{P}_f^n(\mathcal{B}_\mu) \leq Cn^{-p}.$$

Par l'inégalité (5.40), le fait que  $l_j \geq l_{j_1} \asymp (\log n)^{p/2}$  pour  $j \in \{j_1, \dots, j_2\}$  et le Lemme 5.7.3, pour  $\mu$  suffisamment grand, on a

$$\mathcal{V} \leq Cn^{-p}.$$

Les majorations obtenues pour  $\mathcal{U}$  et  $\mathcal{V}$  prouvent l'hypothèse (F2). Cela termine la preuve du Théorème 5.6.2  $\square$

PREUVE DU LEMME 5.7.3. Étudions les majorations de  $\mathbb{P}_f^n(\mathcal{B}_\mu)$  et  $\mathbb{P}_f^n(\mathcal{A}_\mu)$ .

- *Majoration de  $\mathbb{P}_f^n(\mathcal{B}_\mu)$ .* Les variables aléatoires

$$(|\psi_{j,k}(X_1)|^2 g(X_1)^{-1}, \dots, |\psi_{j,k}(X_n)|^2 g(X_n)^{-1}),$$

sont i.i.d. Par l'hypothèse de minoration faite sur  $g$ , on a

$$|\psi_{j,k}(X_i)|^2 g(X_i)^{-1} \leq \|1/g\|_\infty \|\psi\|_\infty^2 2^j \quad \text{et} \quad \mathbb{E}_f^n(|\psi_{j,k}(X_1)|^2 g(X_1)^{-1}) = 1.$$

De ce fait, pour tout  $j \in \{j_1, \dots, j_2\}$ , l'inégalité de Hoeffding justifie l'existence d'une constante  $C > 0$  vérifiant

$$\mathbb{P}_f^n(\mathcal{B}_\mu) \leq 2 \exp(-Cn\mu^2 2^{-2j}) \leq 2 \exp(-Cn\mu^2 2^{-2j_2}) \leq 2n^{-C\mu^2}.$$

On obtient le résultat souhaité en prenant  $\mu$  suffisamment grand.

- *Majoration de  $\mathbb{P}_f^n(\mathcal{A}_\mu)$ .* L'objectif est d'appliquer l'inégalité de Talagrand énoncé dans le Lemme 5.7.1. Considérons l'ensemble  $\mathcal{C}_q$  défini par (5.33) ainsi que la classe de fonctions  $\mathcal{F}'$  définie par

$$\mathcal{F}' = \{h; h(x) = g(x)^{-1} \sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} a_{j,k} a_{j,k'} \psi_{j,k}(x) \psi_{j,k'}(x), \quad a \in \mathcal{C}_q\}.$$

## 5.7. DÉMONSTRATIONS

---

On a

$$\begin{aligned} & \sup_{a \in \mathcal{C}_q} \left( \sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} a_{j,k} a_{j,k'} \left( n^{-1} \sum_{i=1}^n g(X_i)^{-1} \psi_{j,k}(X_i) \psi_{j,k'}(X_i) \right) - \sum_{(K)} |a_{j,k}|^2 \right) \\ &= \sup_{h \in \mathcal{F}'} r_n(h), \end{aligned}$$

où  $r_n$  désigne l'opérateur défini dans le Lemme 5.7.1.

Par conséquent, il suffit d'évaluer les paramètres  $T$ ,  $H$  et  $v$  inhérents à l'inégalité de Talagrand.

– *Evaluation de  $T$ .* Soit  $h \in \mathcal{F}'$ . En utilisant l'inégalité de Hölder, l'hypothèse de minoration faite sur  $g$  et la propriété de concentration (2.5), on trouve

$$|h(x)| \leq \|1/g\|_\infty \sum_{(K)} |a_{j,k}|^2 \sum_{(K)} |\psi_{j,k}(x)|^2 \leq C2^j, \quad x \in [0, 1].$$

D'où  $T = C2^j$ .

– *Evaluation de  $H$ .* L'inégalité de Hölder  $l_2$  et l'inégalité de Hölder entraînent

$$\begin{aligned} & \mathbb{E}_f^n \left( \sup_{a \in \mathcal{C}_q} \sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} a_{j,k} a_{j,k'} \left( \sum_{i=1}^n \epsilon_i g(X_i)^{-1} \psi_{j,k}(X_i) \psi_{j,k'}(X_i) \right) \right) \\ & \leq \sup_{a \in \mathcal{C}_q} \left( \sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} |a_{j,k}|^2 |a_{j,k'}|^2 \right)^{1/2} \dots \\ & \quad \left( \sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} \mathbb{E}_f^n \left( \left| \sum_{i=1}^n \epsilon_i (g(X_i)^{-1} \psi_{j,k}(X_i) \psi_{j,k'}(X_i)) \right|^2 \right) \right)^{1/2} \\ & \leq C \left( \sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} \mathbb{E}_f^n \left( \left| \sum_{i=1}^n \epsilon_i (g(X_i)^{-1} \psi_{j,k}(X_i) \psi_{j,k'}(X_i)) \right|^2 \right) \right)^{1/2}. \quad (5.41) \end{aligned}$$

Puisque  $(\epsilon_1, \dots, \epsilon_n)$  sont des variables de Rademacher i.i.d indépendantes des  $(X_1, \dots, X_n) = \mathbb{X}$ , l'inégalité de Khintchine et l'hypothèse de minoration faite sur  $g$  impliquent

$$\begin{aligned} & \mathbb{E}_f^n \left( \left| \sum_{i=1}^n \epsilon_i (g(X_i)^{-1} \psi_{j,k}(X_i) \psi_{j,k'}(X_i)) \right|^2 \right) \\ &= \mathbb{E}_f^n \left( \mathbb{E}_f^n \left( \left| \sum_{i=1}^n \epsilon_i (g(X_i)^{-1} \psi_{j,k}(X_i) \psi_{j,k'}(X_i)) \right|^2 \middle| \mathbb{X} \right) \right) \\ &\leq C \mathbb{E}_f^n \left( \sum_{i=1}^n |g(X_i)|^{-2} |\psi_{j,k}(X_i)|^2 |\psi_{j,k'}(X_i)|^2 \right) \\ &\leq C \|1/g\|_\infty^2 n \mathbb{E}_f^n (|\psi_{j,k}(X_1)|^2 |\psi_{j,k'}(X_1)|^2). \quad (5.42) \end{aligned}$$

La propriété de concentration (2.5) et les inégalités (5.41) et (5.42) impliquent

$$\begin{aligned} & \mathbb{E}_f^n \left( \sup_{a \in \mathcal{C}_q} \sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} a_{j,k} a_{j,k'} \left( \sum_{i=1}^n \epsilon_i g(X_i)^{-1} \psi_{j,k}(X_i) \psi_{j,k'}(X_i) \right) \right) \\ & \leq C(n \mathbb{E}_f^n \left( \sum_{k \in \mathcal{B}_{j,K}} |\psi_{j,k}(X_i)|^2 \right))^1/2 \leq C n^{1/2} 2^j. \end{aligned}$$

D'où  $H = C 2^j n^{-1/2}$ .

– *Evaluation de  $v$ .* En utilisant l'hypothèse de minoration faite sur  $g$ , l'inégalité de Hölder et la propriété de concentration (2.5), on a

$$\begin{aligned} \sup_{h \in \mathcal{F}} \text{Var}(h(X_1)) & \leq \sup_{a \in \mathcal{C}_q} \mathbb{E}_f^n(|g(X_1)|^{-2}) \sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} a_{j,k} a_{j,k'} \psi_{j,k}(X_1) \psi_{j,k'}(X_1)^2 \\ & \leq \|1/g\|_\infty^2 \sup_{a \in \mathcal{C}_q} \left( \sum_{(K)} |a_{j,k}|^2 \right)^2 \mathbb{E}_f^n \left( \left( \sum_{(K)} |\psi_{j,k}(X_1)|^2 \right)^2 \right) \leq C 2^{2j}. \end{aligned}$$

D'où  $v = C 2^{2j}$ .

Si  $t = 2^{-1} \mu$  alors

$$(t^2 v^{-1} \wedge t T^{-1}) \geq C (\mu^2 2^{-2j} \wedge \mu 2^{-j}) = C \mu^2 2^{-2j}.$$

Choisissons un tel  $t$ . Pour tout  $j \in \{j_1, \dots, j_2\}$  et  $\mu$  suffisamment grand, l'inégalité de Talagrand donne

$$\begin{aligned} \mathbb{P}_f^n(\mathcal{A}_\mu) & \leq \mathbb{P}_f^n \left( \sup_{a \in \mathcal{C}_q} \sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} a_{j,k} a_{j,k'} \left( n^{-1} \sum_{i=1}^n g(X_i)^{-1} \psi_{j,k}(X_i) \psi_{j,k'}(X_i) \right) - \dots \right. \\ & \quad \left. \sum_{(K)} |a_{j,k}|^2 \geq 2^{-1} \mu + C 2^j n^{-1/2} \right) \leq \mathbb{P}_f^n \left( \sup_{h \in \mathcal{F}} r_n(h) \geq t + C_2 H \right) \\ & \leq \exp(-n C_1 (t^2 v^{-1} \wedge t T^{-1})) \leq \exp(-n C \mu^2 2^{-2j}) \\ & \leq \exp(-n C \mu^2 2^{-2j_2}) \leq n^{-p}. \end{aligned}$$

Cela termine la preuve du Lemme 5.7.3. □

**Preuve du Théorème 5.6.4.** Pour la preuve des hypothèses (F1) et (F3), voir l'article de Johnstone et al. (2004, Proof of Claim 1). Montrons que l'hypothèse (F2) est satisfaite.

Si  $1 \leq p < 2$  alors une inégalité de norme  $l_p$  implique

$$\begin{aligned} & \mathbb{P}_f^n \left( (l_j^{-1} \sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^p)^{1/p} \geq 2^{-1} \mu 2^{\delta_j} n^{-1/2} \right) \\ & \leq \mathbb{P}_f^n \left( (l_j^{-1} \sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^2)^{1/2} \geq 2^{-1} \mu 2^{\delta_j} n^{-1/2} \right). \end{aligned}$$

## 5.7. DÉMONSTRATIONS

---

Par conséquent, il suffit de montrer l'hypothèse (F2) pour  $p \geq 2$ .

L'objectif est d'appliquer l'inégalité de Cirelson (5.32).

Posons  $\tilde{e}_{j,k} = \hat{\beta}_{j,k} - \beta_{j,k} = n^{-1/2} \sum_{l \in C_j} \mathcal{F}^*(W)(l) \mathcal{F}(g)(l)^{-1} \mathcal{F}(\psi_{j,k}^m)(l)$ . Considérons l'ensemble  $C_q$  défini en (5.33) et le processus gaussien centré  $\tilde{\mathcal{Z}}(a)$  défini par

$$\tilde{\mathcal{Z}}(a) = \sum_{(K)} a_{j,k} \tilde{e}_{j,k}.$$

Par un argument de dualité, on a

$$\sup_{a \in \mathcal{C}_q} \tilde{\mathcal{Z}}(a) = \left( \sum_{(K)} |\tilde{e}_{j,k}|^p \right)^{1/p}.$$

Étudions les valeurs de  $N$  et  $Q$  apparaissant dans l'inégalité de Cirelson (5.32).

– *Evaluation de  $N$ .* L'inégalité de Hölder et l'hypothèse (F1) entraînent

$$\mathbb{E}_f^n(\sup_{a \in \mathcal{C}_q} \tilde{\mathcal{Z}}(a)) = \mathbb{E}_f^n\left(\left(\sum_{(K)} |\tilde{e}_{j,k}|^p\right)^{1/p}\right) \leq \left(\sum_{(K)} \mathbb{E}_f^n(|\tilde{e}_{j,k}|^p)\right)^{1/p} \leq C n^{-1/2} l_j^{1/p} 2^{\delta j}.$$

D'où  $N = C n^{-1/2} l_j^{1/p} 2^{\delta j}$ .

– *Evaluation de  $Q$ .* Remarquons que l'hypothèse (5.4) entraîne

$$|\mathcal{F}(g)(l)|^{-2} \asymp 2^{2\delta j}, \quad l \in C_j.$$

En utilisant le fait que  $\mathcal{F}^*(W)(l) \sim \mathcal{N}(0, 1)$ , l'égalité standard  $\mathbb{E}_f^n(\mathcal{F}^*(W)(l) \overline{\mathcal{F}^*(W)(l')}) = \int_0^1 \exp(-2i\pi(l-l')t) dt = 1$  si  $l = l'$  et 0 sinon, et l'inégalité de Plancherel, il vient

$$\begin{aligned} \sup_{a \in \mathcal{C}_q} \text{Var}_f^n(\tilde{\mathcal{Z}}(a)) &= \sup_{a \in \mathcal{C}_q} \mathbb{E}_f^n\left(\sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} a_{j,k} \tilde{e}_{j,k} \overline{a_{j,k'} \tilde{e}_{j,k'}}\right) \\ &= n^{-1} \sup_{a \in \mathcal{C}_q} \left(\sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} a_{j,k} \overline{a_{j,k'}} \sum_{l \in C_j} \sum_{l' \in C_j} \mathcal{F}(g)(l)^{-1} \mathcal{F}(\psi_{j,k}^m)(l) \right. \\ &\quad \left. \dots (\overline{\mathcal{F}(g)(l')})^{-1} \overline{\mathcal{F}(\psi_{j,k'}^m)(l')} \mathbb{E}_f^n(\mathcal{F}^*(W)(l) \overline{\mathcal{F}^*(W)(l')})\right) \\ &= n^{-1} \sup_{a \in \mathcal{C}_q} \left(\sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} a_{j,k} \overline{a_{j,k'}} \sum_{l \in C_j} |\mathcal{F}(g)(l)|^{-2} \mathcal{F}(\psi_{j,k}^m)(l) \overline{\mathcal{F}(\psi_{j,k'}^m)(l)}\right) \\ &\leq C n^{-1} 2^{2\delta j} \sup_{a \in \mathcal{C}_q} \left(\sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} a_{j,k} \overline{a_{j,k'}} \sum_{l \in C_j} \mathcal{F}(\psi_{j,k}^m)(l) \overline{\mathcal{F}(\psi_{j,k'}^m)(l)}\right) \\ &= C n^{-1} 2^{2\delta j} \sup_{a \in \mathcal{C}_q} \left(\sum_{k \in \mathcal{B}_{j,K}} \sum_{k' \in \mathcal{B}_{j,K}} a_{j,k} \overline{a_{j,k'}} \int_0^1 \psi_{j,k}^m(x) \overline{\psi_{j,k'}^m(x)} dx\right) \\ &= C n^{-1} 2^{2\delta j} \sup_{a \in \mathcal{C}_q} \left(\sum_{k \in \mathcal{B}_{j,K}} |a_{j,k}|^2\right) \leq C 2^{2\delta j} n^{-1}. \end{aligned}$$

D'où  $Q = C2^{2\delta j}n^{-1}$ .

En prenant  $\mu$  suffisamment grand et  $x = 4^{-1}\mu n^{-1/2}l_j^{1/p}2^{\delta j}$ , l'inégalité de Cirelson donne

$$\begin{aligned} \mathbb{P}_f^n((l_j^{-1} \sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^p)^{1/p} \geq 2^{\delta j} 2^{-1} \mu n^{-1/2}) &\leq \mathbb{P}_f^n(\sup_{a \in \mathcal{C}_q} \tilde{Z}(a) \geq x + N) \\ &\leq \exp(-x^2/(2Q)) \leq \exp(-C\mu^2 l_j^{2/p}). \end{aligned}$$

Puisque  $l_j^{2/p} \geq l_{j_1}^{2/p} \asymp \log n$  pour  $j \in \{j_1, \dots, j_2\}$ , en prenant  $\mu$  suffisamment grand on montre l'hypothèse (F2). Cela termine la preuve du Théorème 5.6.4.  $\square$

\* \* \*

# Chapitre 6

## Conclusions et perspectives

*Conclusions des Chapitres 3 et 4.* L'approche minimax sous le risque  $\mathbb{L}^p$  sur des boules de Besov n'est pas bien adaptée à certains problèmes d'estimation fonctionnelle. C'est le cas du modèle de bruit blanc gaussien généralisé et du modèle de régression à pas aléatoires. Les boules de Besov pondérées proposent une alternative satisfaisante. Elles fournissent des vitesses minimax stables pour une large classe de fonctions parasites. De plus, en utilisant les bases d'ondelettes déformées, on peut construire des estimateurs adaptatifs presque optimaux sur ces boules. Outre l'aspect théorique, ces estimateurs se montrent performants en pratique.

*Perspectives.* Il serait intéressant

- de comparer les performances numériques des estimateurs en ondelettes déformées construits par Picard et Kerkycharian (2005) avec d'autres constructions. Nous pensons aux estimateurs par polynômes locaux élaborés par Gaïffas (2005) et ceux reposant sur une base d'ondelettes adaptée aux pas aléatoires développés par Delouille et al. (2004).
- d'adapter les constructions en ondelettes déformées dans un cadre bi-dimensionnel. C'est une direction qui pourrait conduire à des réalisations concrètes dans divers domaines appliqués. Le premier qui nous vient à l'esprit est celui du traitement de l'image.

*Conclusions du Chapitre 5.* Lorsque l'on travaille sous le risque  $\mathbb{L}^p$ , la version  $\mathbb{L}^p$  de l'estimateur BlockShrink a toutes les qualités pour être le meilleur estimateur qui soit. En effet, il est optimal sur une large zone des boules de Besov et possède de remarquables propriétés maxisets. Cela est vrai pour de nombreux modèles statistiques, y compris certains problèmes inverses. Nous avons illustré cette souplesse avec le modèle de bruit blanc gaussien, le modèle de régression à pas aléatoires et le modèle de convolution en bruit blanc gaussien.

*Perspectives.* Il serait intéressant d'étudier les performances des estimateurs de seuillage

par blocs pour l'estimation d'une fonction inconnue émanant

- d'un modèle mettant en jeu des variables dépendantes.
- d'un problème inverse complexe nécessitant l'utilisation de "bases" bien adaptées, pas forcément celles construites via les ondelettes. Nous pensons au problème de Wicksell traité dans l'article de Kerkycharian et al. (2006) via le seuillage dur en "needlets".

\* \* \*

# Références

- ABRAMOVICH, F. and BENJAMINI, Y. (1996). Adaptive thresholding of wavelet coefficients. *Computat. Stat. Data Anal.*, 22 :351–361.
- ABRAMOVICH, F., SAPATINAS, T., and SILVERMAN, B. (1998). Wavelet thresholding via a bayesian approach. *J. R. Statist. Soc. B*, 60 :725–749.
- AUTIN, F. (2005). *Point de vue maxiset en estimation non paramétrique*. PhD Thesis, Laboratoire de probabilités et modèles aléatoires, Université Paris VII.
- AUTIN, F. (2006). Maxisets for  $\nu$ -thresholding rules. *A paraître dans TEST*.
- BIRGÉ, L. (1983). Approximation dans les espaces métriques et théorie de l'approximation. *Z. Wahrsch. Ver. Geb.*, 65 :181–237.
- BIRGÉ, L. and MASSART, P. (1997). From model selection to adaptive estimation. *Festschrift for Lucien Le Cam*.
- BRETAGNOLLE, J. and CAROL-HUBER, C. (1979). Estimation des densités : risque minimax. *Z. Wahrsch. Ver. Geb.*, 47 :119–137.
- BROWN, L. and LOW, M. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Annals of statistics*, 24 :2384–2398.
- CAI, T. (1996). Minimax wavelet estimation via block thresholding. *Technical Report, 96-41, Department of Statistics, Purdue University*.
- CAI, T. (1997). On adaptivity of blockshrink wavelet estimator over Besov spaces. *Technical Report, 97-05, Department of Statistics, Purdue University*.
- CAI, T. (1999). Adaptive wavelet estimation : a block thresholding and oracle inequality approach. *Ann. Stat.*, 27 :898–924.
- CAI, T. (2002a). On adaptive wavelet estimation of a derivative and other related linear inverse problems. *J. Statistical Planning and Inference*, 108 :329–349.



- CAI, T. (2002b). On block thresholding in wavelet regression : adaptivity, block size and threshold level. *Statist. Sinica*, 12(4) :1241–1273.
- CAI, T. and BROWN, L. (1999). Wavelet estimation for samples with random uniform design. *Statistics and Probability Letters*, 42 :313–321.
- CAI, T. and CHICKEN, E. (2005). Block thresholding for density estimation : local and global adaptivity. *Journal of Multivariate Analysis*, 95 :76–106.
- CAI, T., LOW, M., and ZHAO, L. (2005). Tradeoffs between global and local risks in nonparametric function estimation. *Technical Report*.
- CAI, T. and SILVERMAN, B. (2001). Incorporating information on neighboring coefficients into wavelet estimation. *Sankhya*, (63) :127–148.
- CAVALIER, L. and TSYBAKOV, A. (2001). Penalized blockwise Stein’s method, monotone oracles and sharp adaptive estimation. *Math. Meth. Statist.*, 10(3) :247–282.
- CAVALIER, L. and TSYBAKOV, A. (2002). Sharp adaptation for inverse problems with random noise. *Probability Theory and Related Fields*, 123 :323–354.
- CHESNEAU, C. (2005). A maxiset approach of a gaussian noise model. *A paraître dans TEST*.
- CHESNEAU, C. (2006a). On wavelet block thresholding estimator : minimax and maxisets approaches under the lp risk. *Preprint Laboratoire de Probabilités et Modèles Aléatoires, Paris VI*.
- CHESNEAU, C. (2006b). Regression in random design and wavelet block thresholding estimator. *Preprint Laboratoire de Probabilités et Modèles Aléatoires, Paris VI*.
- CHESNEAU, C. (2007). Regression in random design : a minimax study. *Statistics and Probability Letters*, 77 :40–53.
- CHICKEN, E. (2003a). Block thresholding and wavelet estimation for nonequispaced samples. *J. Stat. Plan. Inference*, (116) :113–129.
- CHICKEN, E. (2003b). Nonparametric regression on random processes and design. *Florida State University Department of Statistics, Technical Report*.
- CHIPMAN, H. A., KOLACZYK, E., and MCCULLOCH, R. (1997). Adaptive bayesian wavelet shrinkage. *J. Am. Statist. Ass.*, 92 :1413–1421.
- CIRELSON, B., IBRAGIMOV, I., and SUDAKOV, V. (1976). *Norm of Gaussian sample functions*. Berlin, springer verlag ed.
- COHEN, A., DAUBECHIES, I., JAWERTH, B., and VIAL, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, 24(1) :54–81.

- COHEN, A., DEVORE, R., and HOCHMUTH, R. (2000a). Restricted nonlinear approximation. *Constr. Approx*, 16 :85–113.
- COHEN, A., DEVORE, R., KERKYACHARIAN, G., and PICARD, D. (2000b). Maximal spaces with given rate of convergence for thresholding algorithms. *Applied and Computational Harmonic Analysis*, 11(1) :167–191.
- COIFMAN, R. and FEFFERMAN, C. (1974). Weighted norm inequalities for maximal functions and singular integrals. *Studia Math*, 51 :241–250.
- DELOUILLE, V., FRANKE, J., and VON SACHS, R. (2001). Nonparametric stochastic regression with design-adapted wavelets. *Sankhyā Ser. A*, 63 :328–366.
- DELOUILLE, V., SIMOENS, J., and VON SACHS, R. (2004). Smooth design-adapted wavelets for nonparametric stochastic regression. *J. Amer. Statist. Assoc.*, 99 :643–658.
- DELYON, B. and JUDITSKY, A. (1996). On minimax wavelet estimators. *Applied Computational Harmonic Analysis*, 3 :215–228.
- DEVORE, R. and POPOV, V. (1988). Interpolation of Besov spaces. *Trans. Amer. Math. Soc.*, 305 :397–414.
- DONOHO, D. (1995). Nonlinear solutions of linear inverse problems by wavelet-vaguelette decomposition. *Journal of Applied and Computational Harmonic Analysis*, 2 :101–126.
- DONOHO, D. and JOHNSTONE, I. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81 :425–455.
- DONOHO, D. and JOHNSTONE, I. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432) :1200–1224.
- DONOHO, D., JOHNSTONE, I., KERKYACHARIAN, G., and PICARD, D. (1995). Wavelet shrinkage : Asymptotia? *J. Royal Statist. Soc. Ser. B.*, 57 :301–369.
- DONOHO, D., JOHNSTONE, I., KERKYACHARIAN, G., and PICARD, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.*, 24(2) :508–539.
- DONOHO, D. and LOW, M. (1992). Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.*, 20 :944–970.
- DOUKHAN, P. and LEON, J. (1990). Deviation quadratique d’estimateur de densité par projection orthogonale. *Comptes Rendus Acad. Sciences Paris (A)*, 310 :424–430.
- EFROIMOVICH, S. and PINSKER, M. (1996). Sharp-optimal and adaptive estimation for heteroscedastic nonparametric regression. *Statistica Sinica*, 6 :925–242.

- EFROIMOVICH, S. Y. (1985). Nonparametric estimation of a density of unknown smoothness. *Theory Prob. Appl.*, 30 :557–561.
- EFROIMOVICH, S. Y. and PINSKER, M. (1981). Estimation of square-integrable [spectral] density based on a sequence of observations. *Problemy Peredatsii Informatsii*, 17 :50–68.
- EFROIMOVICH, S. Y. and PINSKER, M. (1984). Learning algorithm for nonparametric filtering. *Automation and Remote Control*, 11 :1434–1440.
- ERMAKOV, M. (1989). Minimax estimation of the solution of an ill-posed convolution type problem. *Problems of Information Transmission*, 25 :191–200.
- FAN, A. (1997). Comments on "wavelets in statistics : A review," by a. antoniadis. j. italian. statist. soc. 6. pp. 1348–1360.
- FARELL, R. (1967). On the lack of uniformly consistent sequence of estimates of a density function in certain cases. *Ann. Math. Statist.*, 38 :471–474.
- GAÏFFAS, S. (2005). Régression non-paramétrique et information spatialement inhomogène. *Thèse de l'université Paris VII*.
- GAO, H. (1998). Wavelet shrinkage denoising using the nonnegative garrote. *J. Comput. Graph. Statist.*, 7 :469–488.
- GAO, H. and BRUCE, A. (1996). Waveshrink with firm shrinkage. *Technical report 39, StatSci Division of Mathsoft, Inc.*
- GARCIA-CUERVA, J. and MARTELL, B. (2001). Wavelet characterization of weighted spaces. *Journal of Geometric Analysis*, 11(2) :241–264.
- GOLUBEV, L. (1987). Adaptive asymptotically minimax estimate of smooth signals. *Problems of Info. Trans.*, 23 :57–67.
- HALL, P., KERKYACHARIAN, G., and PICARD, D. (1998). Block thresholding rules for curve estimation using kernel and wavelet methods. *Ann. Statist.*, 26 :942–962.
- HALL, P., KERKYACHARIAN, G., and PICARD, D. (1999). On the minimax optimality of block thresholded wavelet estimators. *Statist. Sinica*, 9(1) :33–49.
- HÄRDLE, W., KERKYACHARIAN, G., PICARD, D., and TSYBAKOV, A. (1998). *Wavelet, Approximation and Statistical Applications*, vol. 129 of *Lectures Notes in Statistics*. Springer Verlag, New York.
- IBRAGIMOV, I. and HASMINSKII, R. (1977). On the estimation of an infinite dimensional parameter in gaussian white noise. *Soviet Math. Dokl.*, 23(5) :1053–1055.
- IBRAGIMOV, I. and HASMINSKII, R. (1982). Bounds for the quality of nonparametric estimation of regression. *Theory Probab. Appl.*, 27(1) :81–94.

- JANSEN, M. (2001). *Noise reduction by wavelet thresholding*, vol. 161. Springer-Verlag, New York, lecture notes in statistics ed.
- JOHNSTONE, I. (1999). Wavelet shrinkage for correlated data and inverse problems : adaptivity results.
- JOHNSTONE, I., KERKYACHARIAN, G., PICARD, D., and RAIMONDO, M. (2004). Wavelet deconvolution in a periodic setting. *Journal of the Royal Statistical Society. Series B. Methodological*, 66 :547–573.
- JUDITSKY, A. (1997). Wavelet estimators : adapting to unknown smoothness. *Math. Methods of Statistics*, (1) :1–20.
- KERKYACHARIAN, G. and PICARD, D. (1993). Density estimation by kernel and wavelets methods : Optimality of Besov spaces. *Statist. Probab. Letters*, 18 :327–336.
- KERKYACHARIAN, G. and PICARD, D. (2000). Thresholding algorithms, maxisets and well concentrated bases. *Test*, 9(2) :283–345.
- KERKYACHARIAN, G. and PICARD, D. (2002). Minimax or maxisets? *Bernoulli*, 8(2) :219–253.
- KERKYACHARIAN, G. and PICARD, D. (2003). Non-linear approximation and muckenhoupt weights. *Constructive Approximation*, 24(2) :123–156.
- KERKYACHARIAN, G. and PICARD, D. (2005). Regression in random design and warped wavelets. *Bernoulli*, 10(6) :1053–1105.
- KERKYACHARIAN, G., PICARD, D., PENCHO, P., and WILLER, T. (2006). Need-  
vd : a second generation wavelet algorithm for estimation in inverse problems. *Preprint, Université Paris VII, LPMA*.
- KERKYACHARIAN, G., PICARD, D., and RAIMONDO, M. (2005). Adaptive boxcar deconvolution on full lebesgue measure sets. Technical report, Department of Statistics.
- KERKYACHARIAN, G., PICARD, D., and TRIBOULEY, K. (1996). Lp adaptive density estimation. *Bernoulli*, 2 :229–247.
- KOO, J. (1993). Optimal rates of convergence for nonparametric statistical inverse problems. *Ann. Statist.*, 21 :590–599.
- KOROSTELEV, A. and TSYBAKOV, A. B. (1993). *Minimax Theory of Image Reconstruction*. Springer-Verlag, New York.
- LE CAM, L. and YANG, G. L. (1990). *Asymptotics in Statistics : Some basic concepts..* Springer-Verlag, New York.

- LEPSKII, O. (1990). On a problem of adaptive estimation in gaussian white noise. *Theory Probab. Appl.*, 35 :454–466.
- LEPSKII, O. (1991). Asymptotically minimax adaptive estimation i : Upper bounds. optimally adaptive estimates. *Theory Probab. Appl.*, 36 :682–697.
- LEPSKII, O., MAMMEN, E., and SPOKOINY, V. (1997). Optimal spatial adaptation to inhomogeneous smoothness : An approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.*, 25 :929–947.
- LI, L. and XIAO, Y. (2004). On the minimax optimality of block thresholded wavelet estimators with long memory data. *Preprint*.
- MALLAT, S. (2001). *Une exploration des signaux en ondelettes*. Ecole Polytechnique, ellipses ed.
- MEYER, Y. (1990). *Ondelettes et Opérateurs*. Hermann, Paris.
- MUCKENHOUPT, B. (1972). Weighted norm inequalities for the hardy maximal function. *Transactions of the American Mathematical Society*, pp. 207–226.
- NASON, G. (1996). Wavelet shrinkage using cross-validation. *J. R. Statist. Soc. B*, 58 :463–479.
- PENSKY, M. (1999). Estimation of a smooth density using meyer-type wavelets. *Statist. Decisions*, 17 :111–123.
- PENSKY, M. and VIDAKOVIC, B. (1998). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Stat.*, 27 :2033–2053.
- PENSKY, M. and VIDAKOVIC, B. (2001). On non-equally spaced wavelet regression. *Ann. Inst. Statist. Math*, 53 :681–690.
- PICARD, D. and TRIBOULEY, K. (2000). Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.*, 28(1) :298–335.
- PINSKER, M. (1980). Optimal filtering of square integrable signals in gaussian white noise. *Problems of Info. Trans.*, 16 :120–133.
- QUI, B. (1982). Weighted Besov and Triebel spaces : Interpolation by the real method. *Hiroshima Mathematical Journal*, pp. 581–605.
- RIVOIRARD, V. (2004). Estimation bayésienne non paramétrique. *Thèse de L’université Paris VII*.
- SACKS, J. and YLVISAKER, D. (1981). Asymptotically optimum kernels for density estimation at a point. *Ann. Stat.*, 9(2) :334–346.
- SPECKMAN, P. (1979). Minimax estimates of a linear functionals on a hilbert space. *Manuscript*.

- STEIN, E. (1993). *Harmonic Analysis : Real Variable Methods, Orthogonality, and Oscillatory Integrals*. Princeton University Press.
- STONE, C. (1982). Optimal global rates of convergence for nonparametric regression. *Annals of statistics*, 10(4) :1040–1053.
- TALAGRAND, M. (1994). Sharper bounds for gaussian and empirical processes. *Ann. Probab.*, 22 :28–76.
- TEMLYAKOV, V. (1997). The best m-term approximation and greedy algorithms. *Advances in Comp. Math*, 8 :249–265.
- TSYBAKOV, A. (2004). *Introduction à l'estimation nonparamétrique*. Springer Verlag.
- WALTER, G. G. (1992). Approximation of delta function by wavelets. *Journal of Approximation Theory*, 71 :329–343.
- WALTER, G. G. (1994). *Wavelets and other orthogonal systems in applications*. CRC Press, Boca Raton.
- WEYRICH, N. and WARHOLA, G. (1998). Wavelet shrinkage and generalized cross-validation for image denoising. *IEEE Trans. Im. Proc.*, 7 :82–90.
- WOLFOWITZ, J. (1950). Minimax estimation of the mean of a normal distribution with known variance. *Ann. Math. Statist.*, 21 :218–230.



# Résumé

Nous présentons quelques contributions à l'estimation fonctionnelle par méthodes d'ondelettes. Deux axes de recherches orientent notre travail. *Premier axe : étude de modèles statistiques complexes.* Le point de départ de notre étude est le modèle de bruit blanc gaussien généralisé et le modèle de régression à pas aléatoires. Ceux-ci font intervenir une fonction perturbant l'estimation de la fonction inconnue. Notre objectif est de montrer l'influence exacte de cette fonction parasite via l'approche minimax sous le risque  $\mathbb{L}^p$ . Dans un premier temps, nous utilisons des méthodes en ondelettes pour cerner les limites de cette approche lorsque l'on se place sur des boules de Besov standards. Dans un deuxième temps, nous étudions l'alternative des boules de Besov pondérées et des méthodes en ondelettes déformées. *Deuxième axe : estimation adaptative.* Nous étudions les performances de plusieurs estimateurs de seuillage par blocs en ondelettes sous le risque  $\mathbb{L}^p$ . Nous montrons leurs excellentes propriétés minimax et maxisets pour un large panel de modèles statistiques. En guise d'applications, nous traitons le modèle de régression à pas aléatoires et le modèle de convolution en bruit blanc gaussien.

*Mots clés :* Estimation fonctionnelle, risque  $\mathbb{L}^p$ , bases d'ondelettes déformées, seuillage par blocs, régression à pas aléatoires, convolution en bruit blanc gaussien.

\*                      \*                      \*